Chapter 4

Generalization to Multi-Dimensions and to Systems of Equations

4.1 Multi-Dimensional Scalar Advection-Diffusion Equation

4.1.1 Problem Statement

In the multi-dimensional case, the advection-diffusion equation can be written as follows:

$$\phi_{,t} + \mathbf{u} \cdot \nabla\phi = \nabla \cdot \kappa\nabla\phi + f \tag{4.1}$$

where $\phi = \phi(\mathbf{x}, t)$ is the scalar unknown variable, $\mathbf{x} = \{x_i\}_{1 \le i \le d}$ is the position vector (see Fig. 4.1), $\mathbf{u} = \{u_i\}_{1 \le i \le d}$ is the velocity vector and $f = f(\mathbf{x}, t)$ is the source term.

Moreover, we have introduced the following notations:

- $\nabla\phi = \{\phi_{,x_i}\} = \{\phi_{,i}\}$ is the gradient of ϕ.
- $\mathbf{u} \cdot \nabla\phi = \sum\limits_{i=1}^{d} u_i\phi_{,i}$ is the dot product of $\mathbf{u}$ and $\nabla\phi$. To simplify notation, we will use the summation convention subsequently, i.e., $\mathbf{u} \cdot \nabla\phi = u_i\phi_{,i}$.

The domain $\mathsf{Q} = \Omega\times\,]0, T[$ is a "space-time cylinder" having the generating axis parallel to the time axis (see Fig. 4.2).

We will assume the following properties for the velocity vector and the diffusivity:

- $\mathbf{u} = \mathbf{u}(\mathbf{x})$ (assuming $\mathbf{u} = \text{constant}$ is too simple) is solenoidal ("incompressible"), i.e,

$$0 = \operatorname{div}\mathbf{u} = \nabla \cdot \mathbf{u} = u_{i,i} \tag{4.2}$$

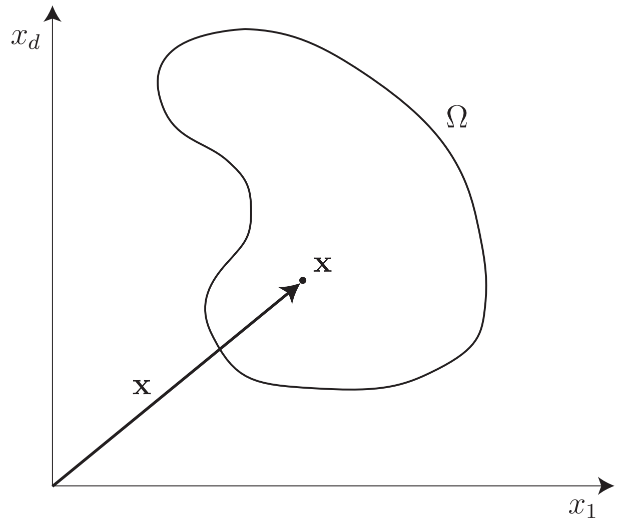


Figure 4.1: Representation of a position vector.

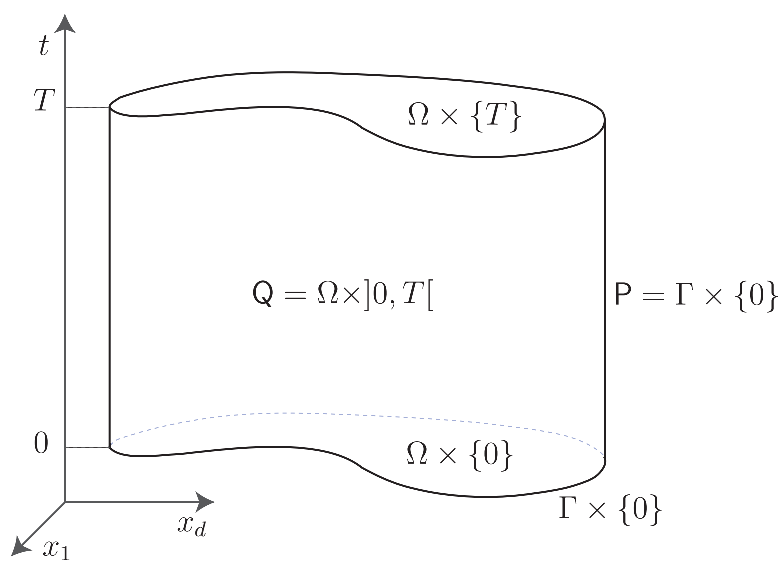


Figure 4.2: Space-time domain.

- The diffusivity matrix

$$\boldsymbol{\kappa} = [\kappa_{ij}] = \begin{bmatrix} \kappa_{11} & \cdots & \kappa_{1d} \\ \vdots & & \vdots \\ \kappa_{d1} & \cdots & \kappa_{dd} \end{bmatrix} \tag{4.3}$$

  is at least symmetric positive semidefinite, but this is an *overkill* in the present context. We will keep the problem simple and assume $\kappa_{ij} = \kappa\delta_{ij}$ where

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \tag{4.4}$$

  is the Kronecker delta. Therefore, we assume that the medium is isotropic, i.e.,

$$\boldsymbol{\kappa} = \kappa \mathbf{I} \tag{4.5}$$

  where κ is a positive constant. Consequently, we have

$$\begin{aligned} \nabla \cdot \boldsymbol{\kappa} \nabla \phi = (\kappa_{ij}\phi_{,j})_{,i} &= \kappa_{ij}\phi_{,ij} \\ &= \kappa\delta_{ij}\phi_{,ij} \\ &= \kappa\phi_{,ii} \\ &= \kappa\Delta\phi \end{aligned} \tag{4.6}$$

The boundary will be chosen to be simple Dirichlet boundary conditions, i.e., defining a function $g : \Upsilon \to \mathbb{R}$, we have $\phi(\mathbf{x}, t) = g(\mathbf{x}, t)$ on $\Upsilon = \Gamma \times]0, T[$. The initial conditions are $\phi(\mathbf{x}, 0) = \bar{\phi}(\mathbf{x})$ on $\Omega \times \{0\}$. Now, we can state the following initial boundary-value problem:

Given f, g, $\mathbf{u}$, κ and $\bar{\phi}$, find ϕ such that the advection-diffusion equation, the boundary conditions and the initial conditions are satisfied.

We have to consider separately the special case of pure advection, i.e., the hyperbolic case with $\kappa = 0$. For this case, the boundary condition specification depends on characteristics (see Fig. 4.3).

On the inflow boundary, we have $\mathbf{u} \cdot \mathbf{n} < 0$. Therefore, we can define the inflow and outflow boundaries as

$$\Gamma^- = \{\mathbf{x} \in \Gamma \mid \mathbf{u}(\mathbf{x}) \cdot \mathbf{n}(\mathbf{x}) < 0\} \tag{4.7}$$

$$\Gamma^+ = \Gamma - \Gamma^- \tag{4.8}$$

Boundary conditions are required on Γ^- but no boundary conditions are permitted on Γ^+. Consequently, for the hyperbolic case, we have to find ϕ such that the advection-diffusion equation (with $\kappa = 0$), the initial conditions and $\phi(\mathbf{x}, t) = g(\mathbf{x}, t)$ on $\Upsilon^- = \Gamma^- \times \,]0, T[$ are satisfied.

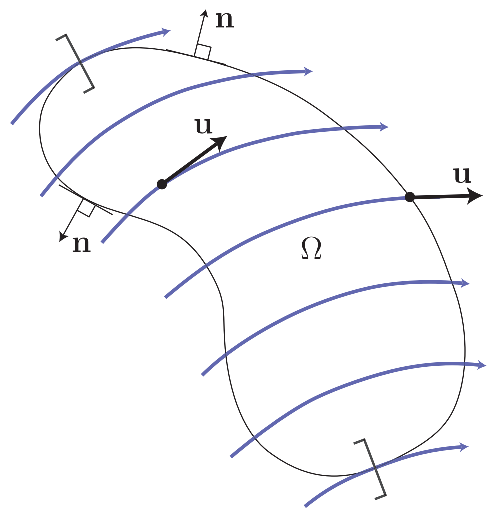


Figure 4.3: Characteristic lines for the hyperbolic problem.

4.1.2 Steady Case

The problem can be stated:

- In the case $\kappa > 0$:

$$\begin{aligned} \mathbf{u} \cdot \nabla \phi &= \kappa \Delta \phi + f \qquad \text{in } \Omega & (4.9) \\ \phi(\mathbf{x}) &= g(\mathbf{x}) \qquad \text{on } \Gamma & (4.10) \end{aligned}$$

- In the case $\kappa = 0$, Γ in (4.10) should be replaced by Γ^-.

Galerkin Method

The finite element spaces are defined by

$$\begin{aligned} \mathcal{V}^h \;&= \big\{ w^h \mid w^h \text{ are } C^0 \text{ finite element functions of order } k\ , \\ &\qquad w^h = 0 \ \text{ on } \Gamma\ (\Gamma^- \text{ for the hyperbolic case}) \big\} & (4.11) \\ \mathcal{S}^h \;&= \big\{ \phi^h \mid \phi^h \text{ are } C^0 \text{ finite element functions of order } k\ , \\ &\qquad \phi^h = g \ \text{ on } \Gamma\ (\Gamma^- \text{ for the hyperbolic case}) \big\} & (4.12) \end{aligned}$$

The Galerkin method can be stated: Find $\phi^h \in \mathcal{S}^h$ such that

$$B(w^h, \phi^h) = L(w^h) \quad \forall w^h \in \mathcal{V}^h \tag{4.13}$$

where

$$\begin{aligned} B(w^h, \phi^h) &\equiv \int_\Omega \left(-\nabla w^h \cdot \mathbf{u}\phi^h + \nabla w^h \cdot \kappa \nabla \phi^h \right) \, d\Omega & (4.14) \\ L(w^h) &\equiv \int_\Omega w^h f \, d\Omega & (4.15) \end{aligned}$$

Before going along, we need to introduce two important results:

Divergence Theorem : For a sufficiently smooth function f (e.g., C^1), we have

$$\int_\Omega f_{,i} \, d\Omega = \int_\Gamma f n_i \, d\Gamma \tag{4.16}$$

Integration-by-parts Formula : For sufficiently smooth functions f and g, we have

$$\int_\Omega f_{,i} g \, d\Omega = \int_\Gamma f g n_i \, d\Gamma - \int_\Omega f g_{,i} \, d\Omega \tag{4.17}$$

Consistency

First, consider the case $\kappa > 0$. We have

$$B(w^h, \phi) - L(w^h) = \int_\Omega \left(-\nabla w^h \cdot \mathbf{u}\phi + \nabla w^h \cdot \kappa \nabla \phi \right) \, d\Omega - \int_\Omega w^h f \, d\Omega \tag{4.18}$$

with

$$\begin{aligned} w^h_{,i} u_i \phi &= (w^h u_i \phi)_{,i} - w^h (u_i \phi)_{,i} \\ &= (w^h u_i \phi)_{,i} - w^h (u_{i,i}\phi + u_i \phi_{,i}) = (w^h u_i \phi)_{,i} - w^h u_i \phi_{,i} & (4.19) \\ w^h_{,i} \kappa \phi_{,i} &= (w^h \kappa \phi_{,i})_{,i} - w^h (\kappa \phi_{,i})_{,i} = (w^h \kappa \phi_{,i})_{,i} - w^h \kappa \phi_{,ii} & (4.20) \end{aligned}$$

Thus,

$$\begin{aligned} B(w^h, \phi) - L(w^h) &= -\int_\Gamma w^h u_i \phi n_i \, d\Gamma + \int_\Omega \left(w^h u_i \phi_{,i} - w^h \kappa \Delta \phi \right) \, d\Omega \\ &\quad + \int_\Gamma w^h \kappa \phi_{,i} n_i \, d\Gamma - \int_\Omega w^h f \, d\Omega \\ &= \int_\Omega w^h \left(\mathbf{u} \cdot \nabla \phi - \kappa \Delta \phi - f \right) \, d\Omega \\ &= 0 & (4.21) \end{aligned}$$

Consequently, $B(w^h, e) = 0$.

The case $\kappa = 0$ deserves special attention. Now, for $w^h \in \mathcal{V}^h$, $w^h = 0$ on Γ^- only. Therefore, we need to add to the weak formulation an integral over the outflow boundary, i.e.,

$$\int_{\Gamma^+} w^h\, \mathbf{u} \cdot \mathbf{n}\, \phi^h \, d\Gamma \tag{4.22}$$

because, when we integrate by parts, we get

$$\int_{\Gamma} w^h\, \mathbf{u} \cdot \mathbf{n}\, \phi^h \, d\Gamma = \int_{\Gamma^+} w^h\, \mathbf{u} \cdot \mathbf{n}\, \phi^h \, d\Gamma \tag{4.23}$$

Consequently, for this case,

$$B(w^h, \phi^h) = -\int_{\Omega} \nabla w^h \cdot \mathbf{u}\, \phi^h \, d\Omega + \int_{\Gamma^+} w^h\, \mathbf{u} \cdot \mathbf{n}\, \phi^h \, d\Gamma \tag{4.24}$$

and the Galerkin formulation can be stated as in eq. (4.13). Adding this term to the formulation yields a consistent method.

Remark 4.1 *We also need this term in the one-dimensional case when $\kappa = 0$ and the boundary condition is released.*

Stability

In the case $\kappa > 0$, we have

$$\begin{aligned} |||w^h|||^2 &= B(w^h, w^h) \\ &= \int_{\Omega} \left(-\nabla w^h \cdot \mathbf{u}\, w^h + \kappa \nabla w^h \cdot \nabla w^h \right) \, d\Omega \end{aligned} \tag{4.25}$$

But

$$\begin{aligned} -\nabla w^h \cdot \mathbf{u} w^h &= -w^h_{,i} u_i w^h \\ &= -\frac{u_i}{2} \left((w^h)^2 \right)_{,i} \\ &= -\frac{1}{2} \left(u_i (w^h)^2 \right)_{,i} \end{aligned} \tag{4.26}$$

Thus,

$$\begin{aligned} |||w^h|||^2 &= -\int_{\Gamma} \frac{1}{2} u_i n_i (w^h)^2 \, d\Gamma + \kappa ||\nabla w^h||^2_{\Omega} \\ &= \kappa ||\nabla w^h||^2_{\Omega} \end{aligned} \tag{4.27}$$

In the case $\kappa = 0$, we have

$$\begin{aligned} |||w^h|||^2 &= B(w^h, w^h) \\ &= -\int_\Omega \nabla w^h \cdot \mathbf{u}\, w^h \, d\Omega + \int_{\Gamma+} w^h \, \mathbf{u} \cdot \mathbf{n} \, w^h \, d\Gamma \\ &= -\int_\Gamma \frac{1}{2} \mathbf{u} \cdot \mathbf{n} \, (w^h)^2 \, d\Gamma + \int_{\Gamma+} \mathbf{u} \cdot \mathbf{n} \, (w^h)^2 \, d\Gamma \\ &= \frac{1}{2} \int_{\Gamma+} \mathbf{u} \cdot \mathbf{n} \, (w^h)^2 \, d\Gamma \end{aligned} \tag{4.28}$$

As we can see, there is little stability with respect to the outflow boundary.

We note for this study that there is a problem for small positive κ, as in the one-dimensional case. The same problem occurs when $\kappa = 0$ because we have

$$\begin{aligned} |||e^h|||^2 &= B(e^h, e - \eta) \\ &= \int_\Omega \nabla e^h \cdot \mathbf{u} \, \eta \, d\Omega + \dots \end{aligned} \tag{4.29}$$

There is no way to hide the integral on the left-hand side. Therefore, we have a lack of stability, as in the one-dimensional case.

Galerkin/Least-Squares Method

The Galerkin/least-squares method can be formally written as

$$B_{\mathrm{GLS}}(w^h, \phi^h) = L_{\mathrm{GLS}}(w^h) \tag{4.30}$$

where

$$B_{\mathrm{GLS}}(w^h, \phi^h) = B(w^h, \phi^h) + (\tau \mathcal{L} w^h, \mathcal{L} \phi^h)_{\Omega'} \tag{4.31}$$

$$L_{\mathrm{GLS}}(w^h) = L(w^h) + (\tau \mathcal{L} w^h, f)_{\Omega'} \tag{4.32}$$

τ is the same as before with h suitably defined for the multi-dimensional case and $\mathcal{L}w \equiv \mathbf{u} \cdot \nabla w - \kappa \Delta w$. An example of a possible definition of h is shown in Figure 4.4.

The consistency proof is obvious at this point. Let us sketch the stability proof:

- We have $B_{\mathrm{GLS}}(w^h, w^h) = |||w^h|||^2 + ||\tau^{1/2} \mathcal{L} w^h||^2_{\Omega'} \equiv |||w^h|||^2_{\mathrm{GLS}}$.

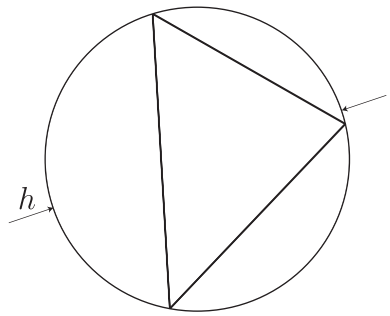


Figure 4.4: Element length in two dimensions.

- Then, in the case $\kappa > 0$, we can establish convergence as follows:

$$\begin{aligned}
|||e^h|||^2_{\text{GLS}} &= B_{\text{GLS}}(e^h, e-\eta) \\
&\leq |-(\nabla e^h \cdot \mathbf{u}, \eta)_\Omega + \kappa(\nabla e^h, \nabla\eta)_\Omega + (\tau^{1/2}\mathcal{L}e^h, \tau^{1/2}\mathcal{L}\eta)_{\Omega'} \\
&\pm (\kappa\Delta e^h, \eta)_{\Omega'}| \\
&\leq \dots \text{ (proceed as usual)} \\
&\leq \frac{\epsilon_1}{2}||\tau^{1/2}\mathcal{L}e^h||^2_{\Omega'} + \frac{1}{2\epsilon_1}||\tau^{-1/2}\eta||^2_\Omega \\
&\quad +\kappa\frac{\epsilon_2}{2}||\Delta e^h||^2_{\Omega'} + \frac{\kappa}{2\epsilon_2}||\eta||^2_\Omega \\
&\quad +\kappa\frac{\epsilon_3}{2}||\nabla e^h||^2_{\Omega} + \frac{\kappa}{2\epsilon_3}||\nabla\eta||^2_\Omega \\
&\quad +\frac{\epsilon_4}{2}||\tau^{1/2}\mathcal{L}e^h||^2_{\Omega'} + \frac{1}{2\epsilon_4}||\tau^{1/2}\mathcal{L}\eta||^2_{\Omega'}
\end{aligned} \tag{4.33}$$

We have the inverse estimate $||\Delta e^h||_{\Omega'} \leq c_I h^{-1}||\nabla e^h||_\Omega$. Therefore,

$$\kappa\frac{\epsilon_2}{2}||\Delta e^h||^2_{\Omega'} \leq \kappa\frac{\epsilon_2}{2}c_I^2 h^{-2}||\nabla e^h||^2_\Omega \tag{4.34}$$

Choosing ϵ_2 such that $\kappa\epsilon_2 c_I^2 h^{-2}/2 = \kappa/4$ yields $\epsilon_2 = c_I^{-2}h^2/2$. Moreover, we choose $\epsilon_1 = \epsilon_3 = \epsilon_4 = 1/2$. Consequently, we get the result

$$|||e^h|||_{\text{GLS}} \leq ch^\ell||\phi||_{k+1} \tag{4.35}$$

with

$$\ell = \begin{cases} k & \text{in the diffusion dominated case,} \\ k+1/2 & \text{in the advection dominated case.} \end{cases} \tag{4.36}$$

Likewise, we get the same result for $|||\eta|||_{\text{GLS}}$ and $|||e|||_{\text{GLS}}$.

- In the case $\kappa = 0$, we get

$$\begin{aligned}
|||e^h|||^2_{\text{GLS}} &\leq |-(\nabla e^h \cdot \mathbf{u}, \eta)_\Omega + ((\mathbf{u}\cdot\mathbf{n})^{1/2} e^h, (\mathbf{u}\cdot\mathbf{n})^{1/2}\eta)_{\Gamma^+} + (\tau^{1/2}\mathcal{L}e^h, \tau^{1/2}\mathcal{L}\eta)_{\Omega'}| \\
&\leq \frac{\epsilon_1}{2}||\tau^{1/2}\mathcal{L}e^h||^2_\Omega + \frac{1}{2\epsilon_1}||\tau^{-1/2}\eta||^2_\Omega \\
&\quad + \frac{\epsilon_2}{2}||(\mathbf{u}\cdot\mathbf{n})^{1/2}e^h||^2_{\Gamma^+} + \frac{1}{2\epsilon_2}||(\mathbf{u}\cdot\mathbf{n})^{1/2}\eta||^2_{\Gamma^+} \\
&\quad + \frac{\epsilon_3}{2}||\tau^{1/2}\mathcal{L}e^h||^2_\Omega + \frac{1}{2\epsilon_3}||\tau^{1/2}\mathcal{L}\eta||^2_\Omega \qquad (4.37)
\end{aligned}$$

We choose $\epsilon_1 = \epsilon_2 = \epsilon_3 = 1/2$. Consequently, we get the result

$$|||e^h|||_{\text{GLS}} \leq ch^{k+1/2}||\phi||_{k+1} \qquad (4.38)$$

Remark 4.2 *We have control over $\mathcal{L}e \equiv \mathbf{u}\cdot\nabla e - \kappa\Delta e$ in the L_2-norm, i.e., control over the streamline derivative only in the multi-dimensional case.*

Remark 4.3 *Convergence can also be established in the following norm (see Franca, Frey and Hughes [78] for details):*

$$|||w^h|||^2_{\text{GLS}} = |||w^h|||^2 + ||\tau^{1/2}\mathbf{u}\cdot\nabla w^h||^2$$

Exercise 4.1 *Show that the SUPG method, namely,*

$$B_{\text{SUPG}}(w^h, \phi^h) = L_{\text{SUPG}}(w^h) \quad \forall w^h \in \mathcal{V}^h \qquad (4.39)$$

where

$$\begin{aligned}
B_{\text{SUPG}}(w^h, \phi^h) &= B(w^h, \phi^h) + (\tau\mathcal{L}^{adv}w^h, \mathcal{L}\phi^h)_{\Omega'} \qquad (4.40)\\
L_{\text{SUPG}}(w^h) &= L(w^h) + (\tau\mathcal{L}^{adv}w^h, f)_{\Omega'} \qquad (4.41)\\
\mathcal{L}^{adv}w^h &= \mathbf{u}\cdot\nabla w^h \qquad (4.42)
\end{aligned}$$

is stable in the sense that

$$\begin{aligned}
B_{\text{SUPG}}(w^h, w^h) &\geq \frac{1}{2}\left(|||w^h|||^2 + ||\tau^{1/2}\mathbf{u}\cdot\nabla w^h||^2_{\Omega'}\right) \\
&\equiv |||w^h|||^2_{\text{SUPG}} \quad \forall w^h \in \mathcal{V}^h \qquad (4.43)
\end{aligned}$$

and that a sufficient condition for stability is that $\tilde{\xi} \leq 2\alpha/C_I^2$, where $\tau = \tilde{\xi}h/(2|\mathbf{u}|)$ element-wise, C_I is an inverse estimate constant, and α is the element Péclet number. Consequently, convergence of SUPG can be established in the norm $|||\cdot|||_{\text{SUPG}}$.

Remark 4.4 *Bochev* et al. *[21] have shown that SUPG actually enjoys better stability properties than GLS. The reason for this concerns the fact that there is an upper bound for τ in the analysis of GLS. Ostensibly, the same thing occurs for SUPG, but in this case there is a transition of the stability from "coercivity," as in (4.43) to "inf-sup" stability. For details of the analysis and numerical results, see [21].*

4.1.3 Time-dependent Case

As in the one-dimensional case, we have to discretize the space-time domain into space-time slabs (see Fig. 4.5), each of which is also discretized in an unstructured way with $(d+1)$-dimensional elements. However, we can think of the simpler situation of elements being structured with respect to time, e.g., elements being bilinear in time and linear in space as depicted in Figure 4.6, or even the simpler case where the elements are constant in time.

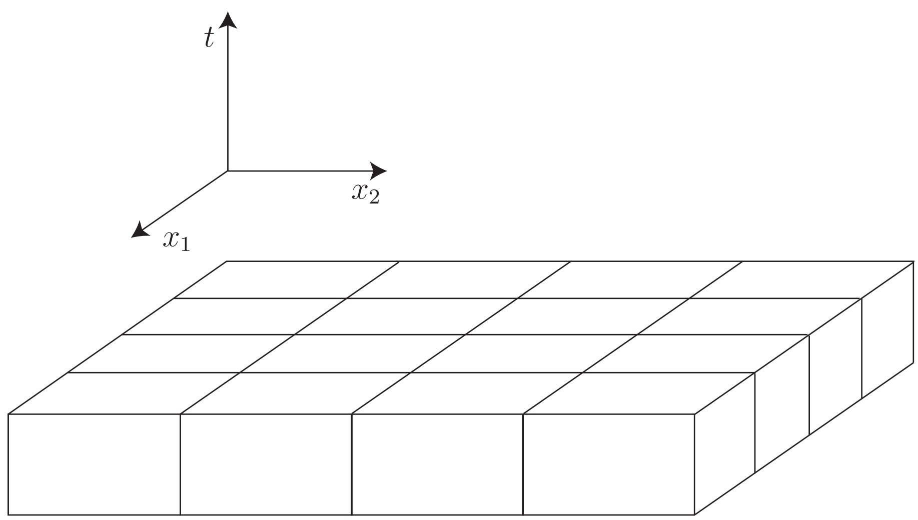


Figure 4.5: Multidimensional space-time slab.

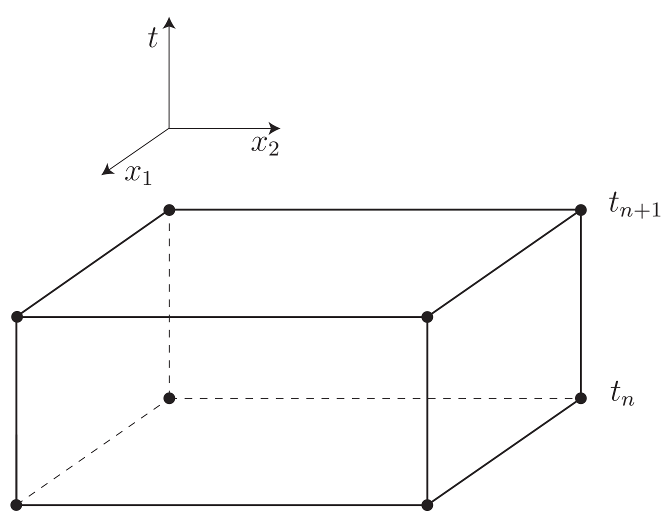


Figure 4.6: Example of a structured space-time element.

Time-discontinuous Galerkin Method

The time-discontinuous Galerkin formulation can be stated as

$$B(w^h, \phi^h)_n = L(w^h)_n \qquad n = 0, 1, \ldots, n_{\max} - 1 \tag{4.44}$$

where

$$\begin{aligned} B(w^h, \phi^h)_n &= \int_{t_n}^{t_{n+1}} \left(-(\dot{w}^h, \phi^h)_\Omega + B(w^h, \phi^h) \right) \, dt \\ &\quad + (w^h(t_{n+1}^-), \phi^h(t_{n+1}^-))_\Omega \end{aligned} \tag{4.45}$$

$$L(w^h)_n = \int_{t_n}^{t_{n+1}} L(w^h) \, dt + (w^h(t_n^+), \phi^h(t_n^-))_\Omega \tag{4.46}$$

This is formally identical to the one-dimensional case.

Galerkin/Least-Squares Method

The Galerkin/least-squares method can be stated as

$$B_{\mathrm{GLS}}(w^h, \phi^h)_n = L_{\mathrm{GLS}}(w^h)_n \qquad n = 0, 1, \ldots, n_{\max} - 1 \tag{4.47}$$

where

$$B_{\mathrm{GLS}}(w^h, \phi^h)_n = B(w^h, \phi^h)_n + (\tau \mathcal{L}_t w^h, \mathcal{L}_t \phi^h)_{\tilde{\mathsf{Q}}_n} \tag{4.48}$$

$$L_{\mathrm{GLS}}(w^h)_n = L(w^h)_n + (\tau \mathcal{L}_t w^h, f)_{\tilde{\mathsf{Q}}_n} \tag{4.49}$$

Analysis of the Methods

As for the steady Galerkin method, we can define

$$\mathbf{B}(w^h, \phi^h) \equiv \sum_{n=0}^{n_{\max}-1} \left(B(w^h, \phi^h)_n - (w^h(t_n^+), \phi^h(t_n^-))_\Omega \right) \tag{4.50}$$

$$\mathbf{L}(w^h) \equiv \sum_{n=0}^{n_{\max}-1} \int_{t_n}^{t_{n+1}} L(w^h) \, dt \tag{4.51}$$

The Galerkin norm has the same definition as in the one-dimensional case, i.e.,

$$\| w^h \|^2 = \mathbf{B}(w^h, w^h) \tag{4.52}$$

For the Galerkin/least-squares method, we have

$$\mathbf{B}_{\mathrm{GLS}}(w^h, \phi^h) \equiv \mathbf{B}(w^h, \phi^h) + (\tau \mathcal{L}_t w^h, \mathcal{L}_t \phi^h)_{\tilde{\mathsf{Q}}} \tag{4.53}$$

$$\mathbf{L}_{\mathrm{GLS}}(w^h) \equiv \mathbf{L}(w^h) + (\tau \mathcal{L}_t w^h, f)_{\tilde{\mathsf{Q}}} \tag{4.54}$$

and the Galerkin/least-squares norm is

$$|\!|\!| w^h |\!|\!|^2_{\mathrm{GLS}} = |\!|\!| w^h |\!|\!|^2 + ||\tau^{1/2}\mathcal{L}_t w^h||^2_{\tilde{\mathsf{Q}}} \tag{4.55}$$

For the Galerkin/least-squares method, we can prove

$$|\!|\!| e |\!|\!|_{\mathrm{GLS}} \leq ch^{\ell}||\phi||_{H^{k+1}(\mathsf{Q})} \tag{4.56}$$

with

$$\ell = \begin{cases} k & \text{in the diffusion dominated case,} \\ k + 1/2 & \text{in the advection dominated case.} \end{cases} \tag{4.57}$$

Exercise 4.2 *State all hypotheses and prove all of the above results.*

Remark 4.5 *We could have developed an SUPG formulation just by dropping the diffusion term from the least-squares operator.*

Remark 4.6 *For the hyperbolic case ($\kappa = 0$) or for spatially linear elements (i.e., linear triangles in two dimensions or linear tetrahedra in three dimensions), the SUPG and Galerkin/least-squares formulations are identical.*

Remark 4.7 *We can now justify the word "streamline" in the definition of SUPG. Consider the steady hyperbolic case and compare the left-hand side operators of Galerkin and Galerkin/least-squares:*

$$\begin{aligned} B_{\mathrm{GLS}}(w^h, \phi^h) - B(w^h, \phi^h) &= (\tau \mathcal{L} w^h, \mathcal{L}\phi^h)_{\Omega'} \\ &= (\tau \mathbf{u} \cdot \nabla w^h, \mathbf{u} \cdot \nabla \phi^h)_{\Omega'} \\ &= \sum_e \int_{\Omega_e} \tau u_i w^h_{,i} u_j \phi^h_{,j} \, d\Omega \\ &= \sum_e \int_{\Omega_e} w^h_{,i} (\tau u_i u_j) \phi^h_{,j} \, d\Omega \\ &= \sum_e \int_{\Omega_e} \nabla w^h \cdot \tau \mathbf{u} \otimes \mathbf{u} \nabla \phi^h \, d\Omega \end{aligned} \tag{4.58}$$

But we have $\kappa = 0$; therefore, $\tau = h/2||\mathbf{u}||$. Thus,

$$\begin{aligned} \tilde{\kappa} &= \tau \mathbf{u} \otimes \mathbf{u} \\ &= \frac{h}{2||\mathbf{u}||} \mathbf{u} \otimes \mathbf{u} \\ &= \frac{h||\mathbf{u}||}{2} \frac{\mathbf{u}}{||\mathbf{u}||} \otimes \frac{\mathbf{u}}{||\mathbf{u}||} \\ &= \tilde{\kappa} \hat{\mathbf{u}} \otimes \hat{\mathbf{u}} \end{aligned} \tag{4.59}$$

In the global coordinate system in three dimensions, we have

$$\hat{\mathbf{u}} \otimes \hat{\mathbf{u}} = \begin{bmatrix} \hat{u}_1{}^2 & \hat{u}_1\hat{u}_2 & \hat{u}_1\hat{u}_3 \\ \hat{u}_2\hat{u}_1 & \hat{u}_2{}^2 & \hat{u}_2\hat{u}_3 \\ \hat{u}_3\hat{u}_1 & \hat{u}_3\hat{u}_2 & \hat{u}_3{}^2 \end{bmatrix} \tag{4.60}$$

and if we orient our coordinates at each point so that the first direction is parallel to $\mathbf{u}$ *(see Fig. 4.7), we get*

$$\hat{\mathbf{u}} \otimes \hat{\mathbf{u}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{4.61}$$

Consequently, the "artificial diffusion" is introduced in the streamline direction but not in the "crosswind" direction.

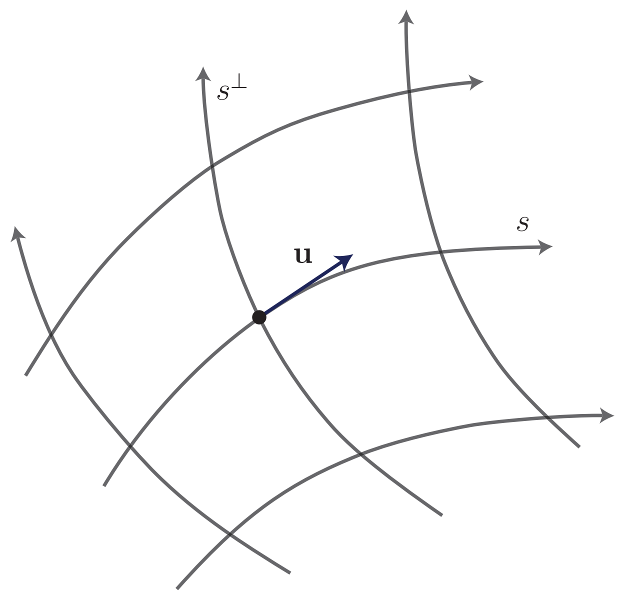


Figure 4.7: Streamline coordinate system.

In multi-dimensions, the SUPG method (which is equivalent to the Galerkin/least-squares method) replicates the optimal one-dimensional method only in the direction of streamline. There is no crosswind diffusion introduced (thus, there is no spurious crosswind diffusion effect). On the other hand, there is no "control" over derivatives orthogonal to the streamlines. In the one-dimensional hyperbolic case, the Galerkin term with the optimal artificial diffusion is equivalent to the classical upwind difference techniques; so, in multi-dimensions, we think of the artificial diffusion as providing "upwinding" along the streamlines (see Fig. 4.8 for the intuitive representation of this idea). Hence, we get the origin of the name "streamline upwind" (Johnson calls this method "streamline diffusion"). However, it should be noted that this interpretation is only for a part of the method and for a very special case. We need to include all effects consistently in the spirit of the Galerkin/least-squares method when the other terms are present. One cannot just add in the diffusivity $\tilde{\kappa}\hat{\mathbf{u}} \otimes \hat{\mathbf{u}}$ *and expect things to work out.*

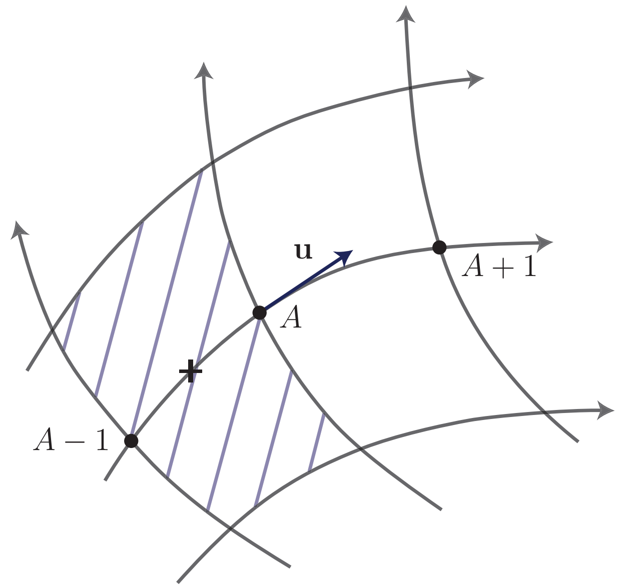


Figure 4.8: Representation of "streamline upwinding". The stencil is centered at A for diffusion whereas it is centered midway between A and $A-1$ for advection.

4.1.4 Numerical Examples

We describe here three numerical examples which first appeared in Brooks and Hughes [39] and are of some historical interest.

Advection of a Cosine Hill in a Rotating Flow Field

The statement of this problem, also known as the "doughnut" problem, is shown in Figure 4.9. The flow consists of a rigid rotation about the center of the bi-unit square domain, with velocity components given by

$$u_1 = -x_2 \qquad , \qquad u_2 = x_1 \tag{4.62}$$

The problem is advection dominated, with a diffusivity of 10^{-6}. Along the external boundary ϕ is set to zero, and on the internal "boundary" OA, ϕ is prescribed to be a cosine hill.

A 30×30 mesh was employed. The exact solution is essentially a pure advection of the OA boundary condition along the circular streamlines. Elevations of ϕ are shown in Figure 4.10.

For this problem, Galerkin and SUPG produce very good results. The SUPG schemes are somewhat better than the Galerkin scheme due to the small-amplitude oscillations of the latter. The effect of pronounced crosswind diffusion is also shown for the Quadrature Upwind Scheme, which is essentially full upwind differencing along coordinate directions.

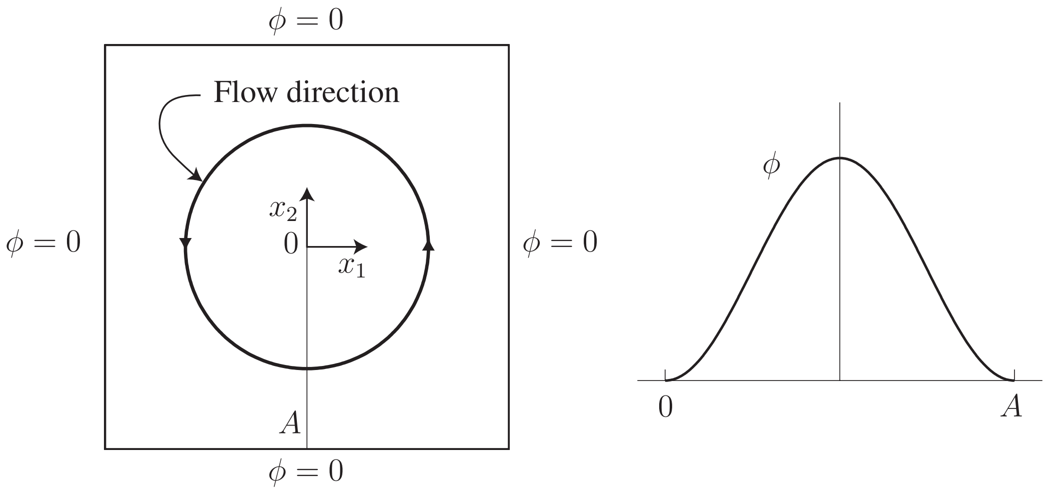
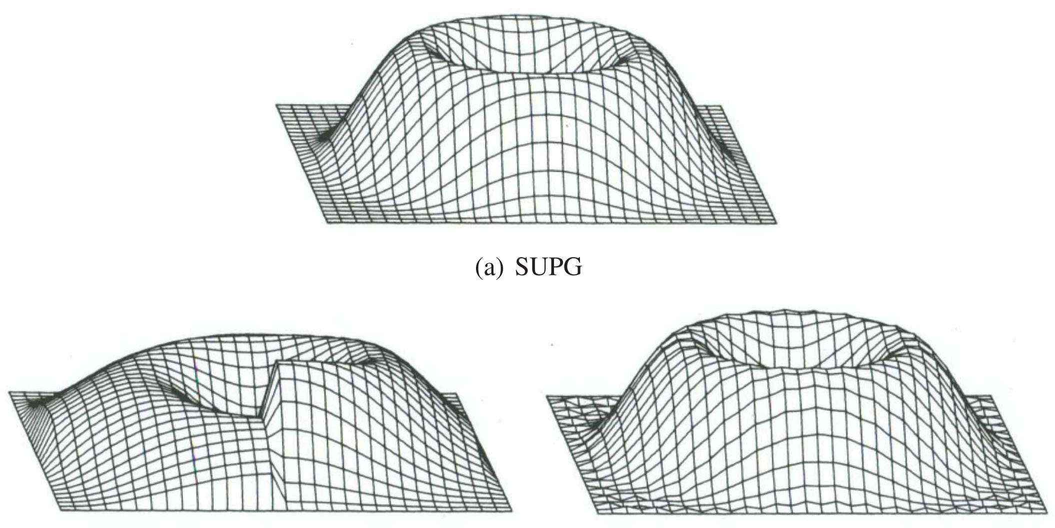


Figure 4.9: Advection in a rotating flow field: problem statement.



(a) SUPG

(b) Quadrature Upwind

(c) Galerkin

Figure 4.10: Advection in a rotating flow field: results.

Rotating Cone

The rotating cone problem has emerged as one of the standard test problems for advection algorithms. The problem statement is shown in Figure 4.11. A mesh of 30×30 was again

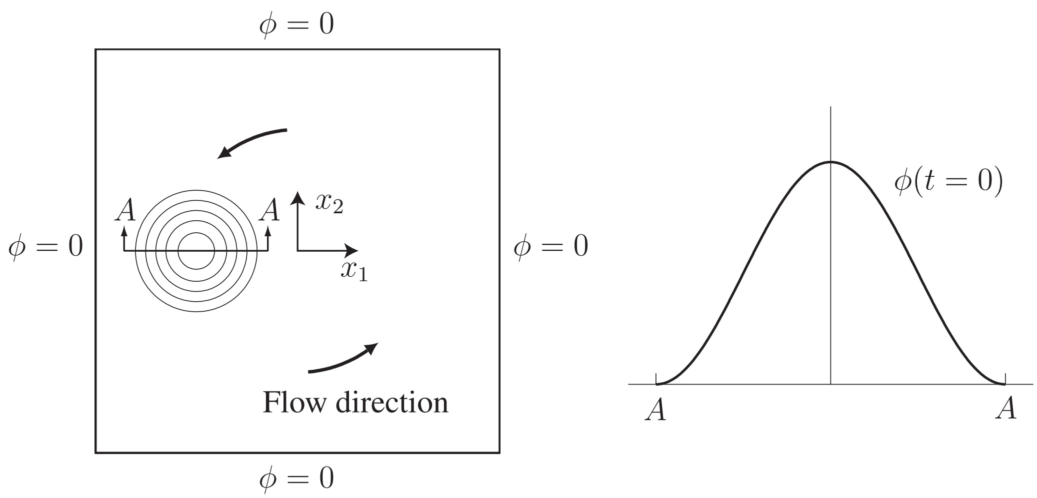
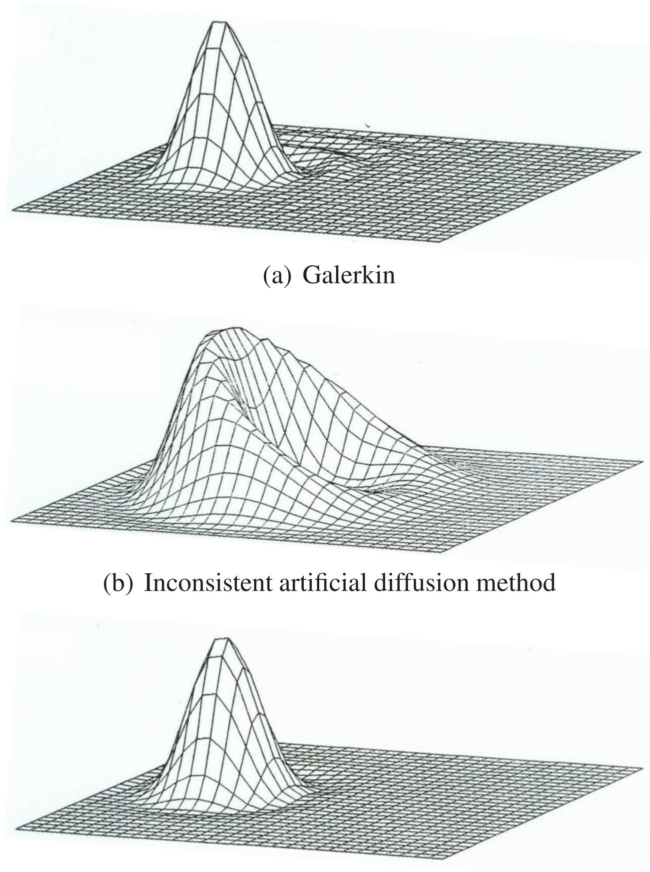


Figure 4.11: Rotating cone: problem statement.

employed. The Courant number at the peak of the cone was approximately $1/4$, corresponding to a full $360°$ rotation of the cone in 200 time steps.

The exact solution consists of a rigid rotation of the cone about the center of the mesh. Essential features of the numerical solutions are phase error, seen as spurious leading and trailing waves, and dissipation error, seen as a reduction in cone height.

Results for a full rotation are shown in Figure 4.12. The inconsistent artificial diffusion method refers to the case in which the streamline diffusion operator is simply added to the Galerkin formulation. The Galerkin method exhibits no dissipation error, but trailing waves of about 5% of the original cone height indicate phase error. The inconsistent artificial diffusion method shows excessive damping of the cone in the direction of its travel. This is not apparent in the figure because all cone heights have been rescaled to their original value by the plotting routine. The SUPG results show much smaller phase error than Galerkin, but the cone is reduced to 88% of its original height. This example once again clearly demonstrates that stabilized finite element schemes should be developed within a consistent weighted-residual formulation, in this case SUPG.



(a) Galerkin

(b) Inconsistent artificial diffusion method

(c) SUPG

Figure 4.12: Rotating cone: results.

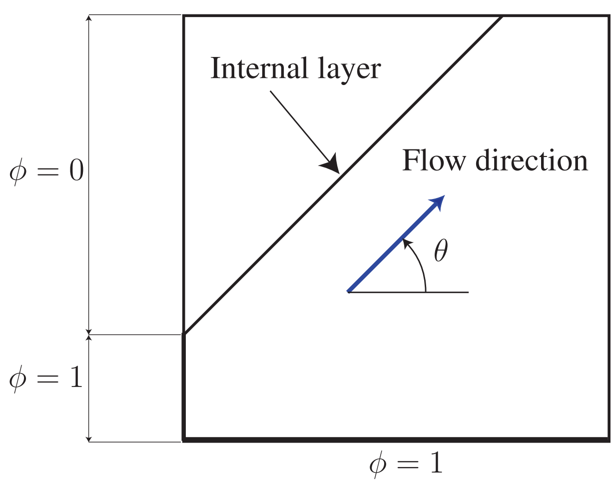


Figure 4.13: Skew advection: problem statement.

Advection Skew to the Mesh

The problem statement is depicted in Figure 4.13. The flow is unidirectional, constant ($||\mathbf{u}|| = 1$), and skew to the mesh. The diffusivity coefficient was taken to be 10^{-6} resulting in a Péclet number of 10^6. In all cases, a 10×10 mesh of equal sized square elements was employed. The inflow boundary condition is discontinuous, as shown, and two different outflow boundary conditions were considered:

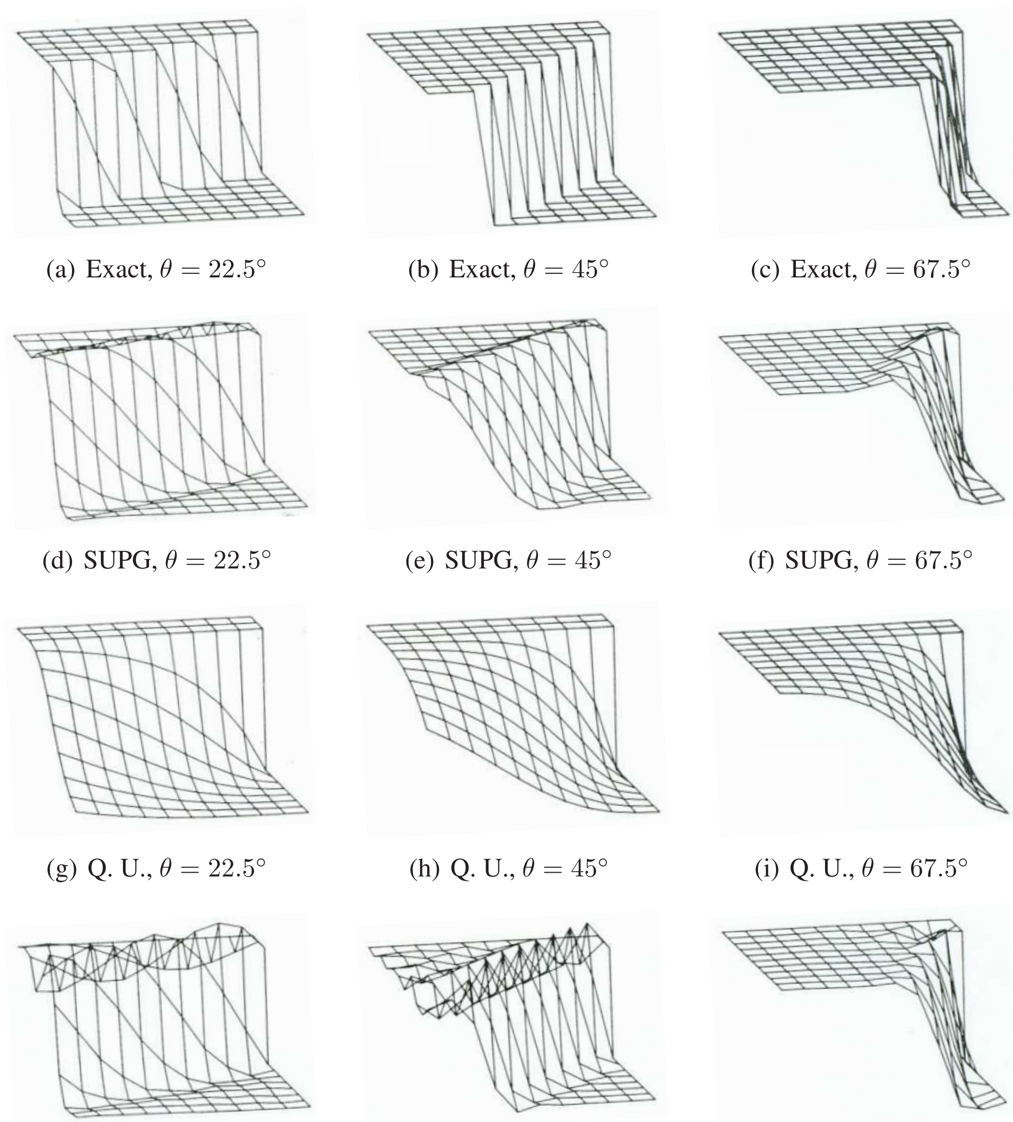
1. *Homogeneous natural boundary conditions*: $\phi_{,n} = 0$.
   For this case, by virtue of the magnitude of the Péclet number, the solution is essentially one of pure advection. The "exact" solution is then simply an advection of the inflow boundary condition in the flow direction. The results shown in Figure 4.14 clearly demonstrate the effect of spurious crosswind diffusion. "Q.U." stands for Quadrature Upwinding. In the advection-dominated case it is essentially the same as full upwinding in each coordinate direction. As may be seen, it exhibits significant spurious crosswind diffusion. Galerkin and SUPG schemes are significantly better in this respect. In passing, we note that when $\theta = 0°$, all cases considered are nodally exact and thus are not shown.

2. *Homogeneous essential boundary conditions*: $\phi = 0$.
   In this case, the solution is identical to the previous one, except in a small neighborhood of the downwind boundary where a very thin "boundary layer" forms. It is a very difficult task for such a crude mesh to capture the essential features of the exact solution under those circumstances. Elevations of ϕ are shown in Figure 4.15. No results are shown for Galerkin, since they are wildly oscillatory and bear no resemblance to the exact results. Again the effect of spurious crosswind diffusion is in evidence. All schemes considered, except Galerkin, are nodally exact for the case $\theta = 0°$ and thus are not shown.

As a conclusion, we can say that the SUPG method is a good linear method with higher-order accurate error estimates when applied to the multi-dimensional advection-diffusion equation, but we would like better performance around discontinuities, and sharp unresolved internal and boundary layers. Linear methods cannot produce *monotone* profiles if they are higher-order accurate (and SUPG with $k \geq 1$ is such a method). This has been a theme of research since shortly after the development of SUPG (see, e.g., Hughes, Mallet and Mizukami [134]), and remains so to the present (see, e.g., Evans, Hughes and Sangalli [72]).

4.2 Multi-Dimensional Advective-Diffusive Systems

This study primarily concerns linear advective-diffusive systems but what is developed initially pertains to nonlinear systems of equations. This is a prelude to the compressible Navier-Stokes equations.



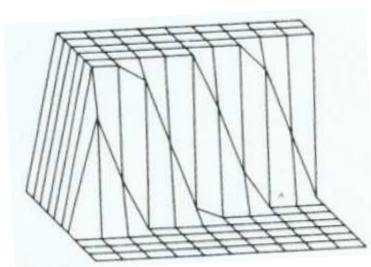
(a) Exact, $\theta = 22.5°$ (b) Exact, $\theta = 45°$ (c) Exact, $\theta = 67.5°$

(d) SUPG, $\theta = 22.5°$ (e) SUPG, $\theta = 45°$ (f) SUPG, $\theta = 67.5°$

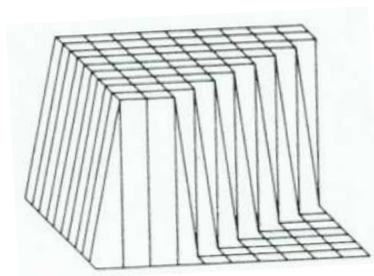
(g) Q. U., $\theta = 22.5°$ (h) Q. U., $\theta = 45°$ (i) Q. U., $\theta = 67.5°$

(j) Galerkin, $\theta = 22.5°$ (k) Galerkin, $\theta = 45°$ (l) Galerkin, $\theta = 67.5°$

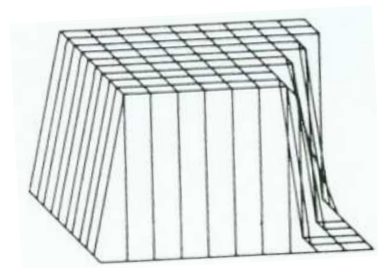
Figure 4.14: Skew advection with homogeneous natural outflow boundary conditions.



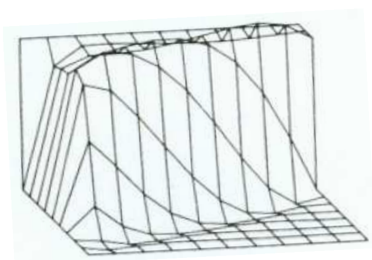
(a) Exact, $\theta = 22.5°$



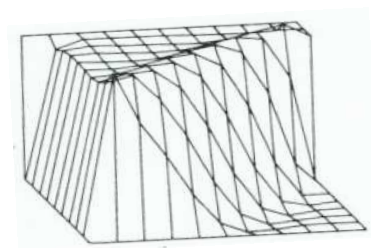
(b) Exact, $\theta = 45°$



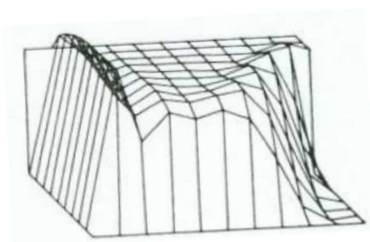
(c) Exact, $\theta = 67.5°$



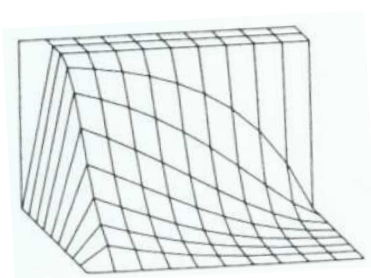
(d) SUPG, $\theta = 22.5°$



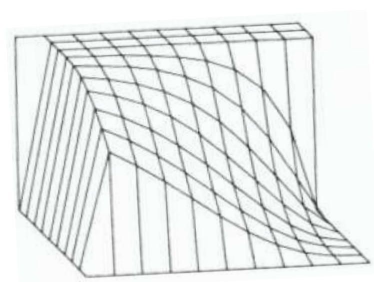
(e) SUPG, $\theta = 45°$



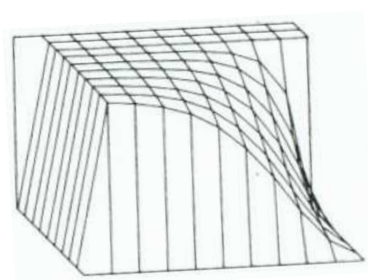
(f) SUPG, $\theta = 67.5°$



(g) Q. U., $\theta = 22.5°$



(h) Q. U., $\theta = 45°$



(i) Q. U., $\theta = 67.5°$

Figure 4.15: Skew advection with homogeneous essential outflow boundary conditions.

4.2.1 Problem Statement

Consider a vector of m unknowns $\mathbf{U} = \mathbf{U}(\mathbf{x}, t)$ where

$$\mathbf{U} = \left\{ \begin{array}{c} U_1 \\ U_2 \\ \vdots \\ U_m \end{array} \right\} \tag{4.63}$$

The system of advection-diffusion equations can be written

$$\mathbf{U}_{,t} + \mathbf{A} \cdot \nabla \mathbf{U} = \nabla \cdot \mathbf{K} \nabla \mathbf{U} + \mathcal{F} \tag{4.64}$$

where

- $\mathbf{U}_{,t} = \dfrac{\partial \mathbf{U}}{\partial t}$.
- $\mathbf{A}^T = [\mathbf{A}_1, \ldots, \mathbf{A}_d]$ is an $m \times (m \cdot d)$ matrix with each $\mathbf{A}_i$ being an $m \times m$ matrix.
- The gradient operator $\nabla^T = \left[\mathbf{I}_m \dfrac{\partial}{\partial x_1}, \ldots, \mathbf{I}_m \dfrac{\partial}{\partial x_d} \right]$ is an $m \times (m \cdot d)$ matrix.
- We have

$$\begin{aligned} \nabla \mathbf{U} &= \left\{ \begin{array}{c} \mathbf{U}_{,1} \\ \mathbf{U}_{,2} \\ \vdots \\ \mathbf{U}_{,d} \end{array} \right\} \\ \mathbf{A} \cdot \nabla \mathbf{U} &= \mathbf{A}^T \nabla \mathbf{U} = \mathbf{A}_1 \mathbf{U}_{,1} + \cdots + \mathbf{A}_d \mathbf{U}_{,d} \end{aligned} \tag{4.65}$$

- The matrix

$$\mathbf{K} = \begin{bmatrix} \mathbf{K}_{11} & \cdots & \mathbf{K}_{1d} \\ \vdots & & \vdots \\ \mathbf{K}_{d1} & \cdots & \mathbf{K}_{dd} \end{bmatrix} \tag{4.66}$$

  is an $(m \cdot d) \times (m \cdot d)$ matrix with each $\mathbf{K}_{ij}$ being an $m \times m$ matrix.

- We have

$$
\begin{aligned}
\mathbf{K}\nabla\mathbf{U} &= \left\{ \begin{array}{c} \mathbf{K}_{11}\mathbf{U}_{,1} + \cdots + \mathbf{K}_{1d}\mathbf{U}_{,d} \\ \vdots \\ \mathbf{K}_{d1}\mathbf{U}_{,1} + \cdots + \mathbf{K}_{dd}\mathbf{U}_{,d} \end{array} \right\} \\
\nabla\cdot(\mathbf{K}\nabla\mathbf{U}) &= \nabla^T(\mathbf{K}\nabla\mathbf{U}) \\
&= (\mathbf{K}_{11}\mathbf{U}_{,1} + \cdots + \mathbf{K}_{1d}\mathbf{U}_{,d})_{,1} \\
&\quad + \cdots \\
&\quad + (\mathbf{K}_{d1}\mathbf{U}_{,1} + \cdots + \mathbf{K}_{dd}\mathbf{U}_{,d})_{,d} \\
&= (\mathbf{K}_{1i}\mathbf{U}_{,i})_{,1} + \cdots + (\mathbf{K}_{di}\mathbf{U}_{,i})_{,d} \\
&= (\mathbf{K}_{ji}\mathbf{U}_{,i})_{,j} \qquad (4.67)
\end{aligned}
$$

- The source vector $\mathcal{F}$ is

$$
\mathcal{F} = \left\{ \begin{array}{c} \mathcal{F}_1 \\ \vdots \\ \mathcal{F}_m \end{array} \right\} \tag{4.68}
$$

Remark 4.8 *The matrices $\mathbf{A}$, $\mathbf{K}$ and $\mathcal{F}$ are algebraic functions of $\mathbf{U}$. Additionally, $\mathcal{F}$ can also depend explicitly on $\mathbf{x}$ and t.*

Remark 4.9 *$\mathbf{A}\cdot\nabla\mathbf{U}$ is a generalized advection term and $\nabla\cdot(\mathbf{K}\nabla\mathbf{U})$ is a generalized diffusion term.*

Remark 4.10 *Assume $\mathbf{K} = 0$. Then (4.64) is called a hyperbolic system if, for all real c_i, $1 \le i \le d$, the matrix $\sum c_i\mathbf{A}_i$ has real eigenvalues and a complete set of eigenvectors. We will see more results later.*

Remark 4.11 *The compressible Navier-Stokes equations can be written in the form of (4.64) in terms of conservative variables $\mathbf{U} = \{\rho,\ \rho\mathbf{u},\ \rho e\}^T$, where ρ is the density, $\mathbf{u}$ is the velocity vector and e is the energy. In the three-dimensional case, $\mathbf{K}$ is a 15×15 matrix and $\mathbf{A}$ a 15×5 matrix. However, this choice of variables does not seem canonical for this case.*

Remark 4.12 *Consider a change of variables $\mathbf{U} \mapsto \mathbf{V}$, $\mathbf{U} = \mathbf{U}(\mathbf{V})$. The system (4.64) becomes:*

$$
\widetilde{\mathbf{A}}_0\mathbf{V}_{,t} + \widetilde{\mathbf{A}}\cdot\nabla\mathbf{V} = \nabla\cdot\widetilde{\mathbf{K}}\nabla\mathbf{V} + \widetilde{\mathcal{F}} \tag{4.69}
$$

with $\widetilde{\mathbf{A}} = \widetilde{\mathbf{A}}(\mathbf{V})$, $\widetilde{\mathbf{K}} = \widetilde{\mathbf{K}}(\mathbf{V})$, $\widetilde{\mathcal{F}} = \mathcal{F}(\mathbf{U}(\mathbf{V}), \mathbf{x}, t)$, $\widetilde{\mathbf{A}}_0 = \dfrac{\partial\mathbf{U}}{\partial\mathbf{V}}$, $\widetilde{\mathbf{A}}^T = [\widetilde{\mathbf{A}}_1, \cdots, \widetilde{\mathbf{A}}_d]$, $\widetilde{\mathbf{A}}_i = \mathbf{A}_i\widetilde{\mathbf{A}}_0$ and

$$
\widetilde{\mathbf{K}} = \begin{bmatrix} \widetilde{\mathbf{K}}_{11} & \cdots & \widetilde{\mathbf{K}}_{1d} \\ \vdots & & \vdots \\ \widetilde{\mathbf{K}}_{d1} & \cdots & \widetilde{\mathbf{K}}_{dd} \end{bmatrix} \tag{4.70}
$$

with $\widetilde{\mathbf{K}}_{ij} = \mathbf{K}_{ij}\widetilde{\mathbf{A}}_0$. The proof is as follows: By the chain rule we get

$$\mathbf{U}_{,t} = \frac{\partial \mathbf{U}}{\partial \mathbf{V}}\mathbf{V}_{,t} = \widetilde{\mathbf{A}}_0\mathbf{V}_{,t}$$

$$\begin{aligned}
\nabla\mathbf{U} = \left\{\begin{array}{c}\mathbf{U}_{,1}\\ \vdots \\ \mathbf{U}_{,d}\end{array}\right\} &= \left\{\begin{array}{c}\frac{\partial \mathbf{U}}{\partial \mathbf{V}}\mathbf{V}_{,1}\\ \vdots \\ \frac{\partial \mathbf{U}}{\partial \mathbf{V}}\mathbf{V}_{,d}\end{array}\right\} \\
&= \left\{\begin{array}{c}\widetilde{\mathbf{A}}_0\mathbf{V}_{,1}\\ \vdots \\ \widetilde{\mathbf{A}}_0\mathbf{V}_{,d}\end{array}\right\} \\
&= \left[\begin{array}{cc}\widetilde{\mathbf{A}}_0 & 0\\ & \ddots \\ 0 & \widetilde{\mathbf{A}}_0\end{array}\right]\left\{\begin{array}{c}\mathbf{V}_{,1}\\ \vdots \\ \mathbf{V}_{,d}\end{array}\right\} \\
&\equiv \left\lceil \widetilde{\mathbf{A}}_0 \right\rfloor \nabla\mathbf{V}
\end{aligned} \tag{4.71}$$

The symbol $\left\lceil \widetilde{\mathbf{A}}_0 \right\rfloor$ represents the block diagonal matrix consisting of d copies of $\widetilde{\mathbf{A}}_0$ along the diagonal.

$$\begin{aligned}
\mathbf{A}\cdot\nabla\mathbf{U} = \mathbf{A}^T\nabla\mathbf{U} &= [\mathbf{A}_1, \cdots, \mathbf{A}_d]\left\lceil \widetilde{\mathbf{A}}_0 \right\rfloor \nabla\mathbf{V} \\
&= \left[\mathbf{A}_1\widetilde{\mathbf{A}}_0, \cdots, \mathbf{A}_d\widetilde{\mathbf{A}}_0\right]\nabla\mathbf{V} \\
&= \left[\widetilde{\mathbf{A}}_1, \cdots, \widetilde{\mathbf{A}}_d\right]\nabla\mathbf{V} \\
&= \widetilde{\mathbf{A}}^T\nabla\mathbf{V} \\
&= \widetilde{\mathbf{A}}\cdot\nabla\mathbf{V}
\end{aligned} \tag{4.72}$$

$$\begin{aligned}
\nabla\cdot\mathbf{K}\nabla\mathbf{U} &= \nabla\cdot\mathbf{K}\left\lceil \widetilde{\mathbf{A}}_0 \right\rfloor \nabla\mathbf{V} \\
&= \nabla\cdot\left[\begin{array}{ccc}\mathbf{K}_{11}\widetilde{\mathbf{A}}_0 & \cdots & \mathbf{K}_{1d}\widetilde{\mathbf{A}}_0\\ \vdots & & \vdots \\ \mathbf{K}_{d1}\widetilde{\mathbf{A}}_0 & \cdots & \mathbf{K}_{dd}\widetilde{\mathbf{A}}_0\end{array}\right]\nabla\mathbf{V} \\
&= \nabla\cdot\left[\begin{array}{ccc}\widetilde{\mathbf{K}}_{11} & \cdots & \widetilde{\mathbf{K}}_{1d}\\ \vdots & & \vdots \\ \widetilde{\mathbf{K}}_{d1} & \cdots & \widetilde{\mathbf{K}}_{dd}\end{array}\right]\nabla\mathbf{V} \\
&= \nabla\cdot\widetilde{\mathbf{K}}\nabla\mathbf{V}
\end{aligned} \tag{4.73}$$

We did not use any linearity property anywhere. Therefore, both (4.64) and (4.69) can be nonlinear equations.

Remark 4.13 *Many physical systems admit changes of variables* $\mathbf{U} \mapsto \mathbf{V}$ *as above such that:*

- $\widetilde{\mathbf{A}}_0$ *is a symmetric positive-definite matrix;*
- *the* $\widetilde{\mathbf{A}}_i$*'s are symmetric matrices;*
- $\widetilde{\mathbf{K}}$ *is a symmetric positive-semidefinite matrix;*

whereas the $\mathbf{A}_i$*'s and* $\mathbf{K}$ *do not satisfy these properties.*

Remark 4.14 *With reference to (4.69) with all the symmetries and definiteness properties of the previous remark, we have:*

1. *If* $\widetilde{\mathbf{K}} = \mathbf{0}$*, (4.69) is called a* ***symmetric hyperbolic system*** *(also called a* ***Friedrichs system****). Hyperbolicity depends on* $\sum_{i=0}^{d} c_i \widetilde{\mathbf{A}}_i$*,* $\forall c_i \in \mathbb{R}$ *having* real *eigenvalues and a complete set of eigenvectors. This is assured since the* $\widetilde{\mathbf{A}}_i$*'s are symmetric and thus have real eigenvalues. The compressible Euler equations written in terms of the entropy variables are in this class of systems.*
2. *If* $\widetilde{\mathbf{K}} > \mathbf{0}$*, (4.69) is called a* ***symmetric parabolic system****.*
3. *If* $\widetilde{\mathbf{K}} \geq \mathbf{0}$ *(but not* $> \mathbf{0}$*), (4.69) is called an* ***incompletely parabolic system****. The compressible Navier-Stokes equations written in terms of the entropy variables are in this class of systems.*

Remark 4.15 *The matrix* $\widetilde{\mathbf{A}}_0$ *plays the role of a metric tensor on* $\mathbb{R}^m$*. If* $\widetilde{\mathbf{A}}_0 = \mathbf{I}_m$*, (4.69) is said to have a* ***Euclidian metric****. If* $\widetilde{\mathbf{A}}_0 \neq \mathbf{I}_m$*, (4.69) is said to have a* ***Riemannian metric****.*

Remark 4.16 *The fact that the Euler and Navier-Stokes equations can be transformed from (4.64) to (4.69) is linked to the existence of an* ***entropy production inequality****.*

Remark 4.17 *In the linear case, all necessary properties to establish stability, consistency and convergence are possessed by (4.69).*

The boundary-value problem can be written

1. *In the case* $\widetilde{\mathbf{K}} > \mathbf{0}$*, Dirichlet boundary conditions are well-posed:*

$$\begin{aligned} \mathbf{V} &= \mathcal{G} \qquad \text{on } \Gamma \\ \mathbf{W} &= \mathbf{0} \qquad \text{on } \Gamma \end{aligned}$$

but this case is not too interesting physically.

2. *In the case* $\widetilde{\mathbf{K}} = \mathbf{0}$*, i.e., the hyperbolic case, assume* $\widetilde{\mathbf{A}}_n = \sum_{i=1}^{d} n_i \widetilde{\mathbf{A}}_i$ *is either* < 0 *or* ≥ 0 *for all* $\mathbf{x}$ *in* Γ*. Therefore, we can define the "inflow" boundary*

$$\Gamma^- = \{\mathbf{x} \in \Gamma \mid \widetilde{\mathbf{A}}_n(\mathbf{x}) < 0\} \tag{4.74}$$

and the outflow boundary

$$\Gamma^+ = \{\mathbf{x} \in \Gamma \mid \widetilde{\mathbf{A}}_n(\mathbf{x}) \geq 0\} \tag{4.75}$$

Consequently, the boundary conditions are

$$\begin{aligned} \mathbf{V} &= \mathcal{G} \qquad \text{on } \Gamma^- \\ \mathbf{W} &= \mathbf{0} \qquad \text{on } \Gamma^- \end{aligned}$$

4.2.2 Steady Case

Variational Form

For the steady case, the variational form can be written

$$B(\mathbf{W}, \mathbf{V}) = L(\mathbf{W}) \tag{4.76}$$

where

$$B(\mathbf{W}, \mathbf{V}) = \int_\Omega \Big(-\nabla\mathbf{W} \cdot \widetilde{\mathbf{A}}\mathbf{V} + \nabla\mathbf{W} \cdot \widetilde{\mathbf{K}}\nabla\mathbf{V}\Big)\, d\Omega \tag{4.77}$$

$$L(\mathbf{W}) = \int_\Omega \mathbf{W} \cdot \widetilde{\mathcal{F}}\, d\Omega \tag{4.78}$$

Consistency

As usual, the consistency proof is based on integrations by parts and derivation of the Euler-Lagrange form of the equations.

$$\begin{aligned} -\nabla\mathbf{W} \cdot \widetilde{\mathbf{A}}\mathbf{V} &= -\mathbf{W}_{,i} \cdot \widetilde{\mathbf{A}}_i\mathbf{V} \\ &= -(\mathbf{W} \cdot \widetilde{\mathbf{A}}_i\mathbf{V})_{,i} + \mathbf{W} \cdot (\widetilde{\mathbf{A}}_i\mathbf{V})_{,i} \\ &= -(\mathbf{W} \cdot \widetilde{\mathbf{A}}_i\mathbf{V})_{,i} + \mathbf{W} \cdot (\widetilde{\mathbf{A}}_{i,i}\mathbf{V} + \widetilde{\mathbf{A}}_i\mathbf{V}_{,i}) \end{aligned} \tag{4.79}$$

Assuming that $\widetilde{\mathbf{A}}_{i,i} \equiv 0$,

$$\begin{aligned} -\int_\Omega \nabla\mathbf{W} \cdot \widetilde{\mathbf{A}}\mathbf{V}\, d\Omega &= -\int_\Gamma (\mathbf{W} \cdot \widetilde{\mathbf{A}}_i\mathbf{V}n_i)\, d\Gamma + \int_\Omega \mathbf{W} \cdot (\widetilde{\mathbf{A}} \cdot \nabla\mathbf{V})\, d\Omega \\ &= -\int_\Gamma \mathbf{W} \cdot \widetilde{\mathbf{A}}_n\mathbf{V}\, d\Gamma + \int_\Omega \mathbf{W} \cdot (\widetilde{\mathbf{A}} \cdot \nabla\mathbf{V})\, d\Omega \end{aligned} \tag{4.80}$$

$$\begin{aligned} \nabla\mathbf{W} \cdot \widetilde{\mathbf{K}}\nabla\mathbf{V} &= \mathbf{W}_{,i} \cdot \widetilde{\mathbf{K}}_{ij}\mathbf{V}_{,j} \\ &= (\mathbf{W} \cdot \widetilde{\mathbf{K}}_{ij}\mathbf{V}_{,j})_{,i} - \Big(\mathbf{W} \cdot (\widetilde{\mathbf{K}}_{ij}\mathbf{V}_{,j})_{,i}\Big) \end{aligned} \tag{4.81}$$

Thus, using the divergence theorem,

$$\int_{\Omega} \nabla \mathbf{W} \cdot \widetilde{\mathbf{K}} \nabla \mathbf{V} \, d\Omega = \int_{\Gamma} \mathbf{W} \cdot (\widetilde{\mathbf{K}}_{ij} \mathbf{V}_{,j} n_i) \, d\Gamma - \int_{\Omega} \mathbf{W} \cdot (\nabla \cdot \widetilde{\mathbf{K}} \nabla \mathbf{V}) \, d\Omega \tag{4.82}$$

In the case $\widetilde{\mathbf{K}} > \mathbf{0}$, with $\mathbf{W} = \mathbf{0}$ on Γ,

$$\begin{aligned} B(\mathbf{W}, \mathbf{V}) - L(\mathbf{W}) &= \int_{\Omega} \mathbf{W} \cdot (\widetilde{\mathbf{A}} \cdot \nabla \mathbf{V} - \nabla \cdot \widetilde{\mathbf{K}} \nabla \mathbf{V} - \widetilde{\mathcal{F}}) \, d\Omega \\ &= \int_{\Omega} \mathbf{W} \cdot (\widetilde{\mathcal{L}} \mathbf{V} - \widetilde{\mathcal{F}}) \, d\Omega \\ &= 0 \end{aligned} \tag{4.83}$$

In the case $\widetilde{\mathbf{K}} = \mathbf{0}$, we proceed as follows. As in the scalar case, an outflow boundary term must be included in order to assure consistency. B is now defined as:

$$B(\mathbf{W}, \mathbf{V}) = \int_{\Omega} -\nabla \mathbf{W} \cdot \widetilde{\mathbf{A}} \mathbf{V} \, d\Omega + \int_{\Gamma^+} \mathbf{W} \cdot \widetilde{\mathbf{A}}_n \mathbf{V} \, d\Gamma \tag{4.84}$$

In the scalar case, the added term was $\int_{\Gamma^+} w u_n \phi \, d\Gamma$. The consistency proof is as follows:

$$\begin{aligned} B(\mathbf{W}, \mathbf{V}) - L(\mathbf{W}) &= \int_{\Omega} -\nabla \mathbf{W} \cdot \widetilde{\mathbf{A}} \mathbf{V} \, d\Omega - \int_{\Omega} \widetilde{\mathcal{F}} \cdot \mathbf{W} \, d\Omega + \int_{\Gamma^+} \mathbf{W} \cdot \widetilde{\mathbf{A}}_n \mathbf{V} \, d\Gamma \\ &= -\int_{\Gamma} \mathbf{W} \cdot \widetilde{\mathbf{A}}_n \mathbf{V} \, d\Gamma + \int_{\Omega} \mathbf{W} \cdot (\widetilde{\mathbf{A}} \cdot \nabla \mathbf{V} - \widetilde{\mathcal{F}}) \, d\Omega + \int_{\Gamma^+} \mathbf{W} \cdot \widetilde{\mathbf{A}}_n \mathbf{V} \, d\Gamma \\ &= -\int_{\Gamma^+} \mathbf{W} \cdot \widetilde{\mathbf{A}}_n \mathbf{V} \, d\Gamma + \int_{\Omega} \mathbf{W} \cdot (\widetilde{\mathcal{L}} \mathbf{V} - \widetilde{\mathcal{F}}) \, d\Omega + \int_{\Gamma^+} \mathbf{W} \cdot \widetilde{\mathbf{A}}_n \mathbf{V} \, d\Gamma \\ &= 0 \end{aligned} \tag{4.85}$$

Stability

In the case $\widetilde{\mathbf{K}} > \mathbf{0}$, the stability statement reads:

$$B(\mathbf{W}, \mathbf{W}) = \int_{\Omega} (-\nabla \mathbf{W} \cdot \widetilde{\mathbf{A}} \mathbf{W} + \nabla \mathbf{W} \cdot \widetilde{\mathbf{K}} \nabla \mathbf{W}) \, d\Omega \tag{4.86}$$

where

$$\begin{aligned} -\nabla \mathbf{W} \cdot \widetilde{\mathbf{A}} \mathbf{W} &= -\mathbf{W}_{,i} \cdot \widetilde{\mathbf{A}}_i \mathbf{W} \\ &= -(\mathbf{W} \cdot \widetilde{\mathbf{A}}_i \mathbf{W})_{,i} + \mathbf{W} \cdot \widetilde{\mathbf{A}}_i \mathbf{W}_{,i} \\ &= -(\mathbf{W} \cdot \widetilde{\mathbf{A}}_i \mathbf{W})_{,i} + \widetilde{\mathbf{A}}_i^T \mathbf{W} \cdot \mathbf{W}_{,i} \\ &= -(\mathbf{W} \cdot \widetilde{\mathbf{A}}_i \mathbf{W})_{,i} + \mathbf{W}_{,i} \cdot \widetilde{\mathbf{A}}_i \mathbf{W} \end{aligned} \tag{4.87}$$

Thus,

$$\int_\Omega \mathbf{W}_{,i} \cdot \widetilde{\mathbf{A}}_i \mathbf{W} \, d\Omega = \frac{1}{2} \int_\Gamma \mathbf{W} \cdot \widetilde{\mathbf{A}}_n \mathbf{W} \, d\Gamma \tag{4.88}$$

and finally,

$$B(\mathbf{W}, \mathbf{W}) = ||\widetilde{\mathbf{K}}^{1/2} \nabla \mathbf{W}||^2_\Omega \tag{4.89}$$

This result is to be compared with the one obtained in the scalar case, viz., $\kappa||\nabla w||^2_\Omega = ||\kappa^{1/2} \nabla w||^2_\Omega$.

In the case $\widetilde{\mathbf{K}} = \mathbf{0}$, we have

$$\begin{aligned} B(\mathbf{W}, \mathbf{W}) &= \int_\Omega -\nabla \mathbf{W} \cdot (\widetilde{\mathbf{A}} \mathbf{W}) \, d\Omega + \int_{\Gamma+} \mathbf{W} \cdot \widetilde{\mathbf{A}}_n \mathbf{W} \, d\Gamma \\ &= \int_\Omega -\mathbf{W}_{,i} \cdot \widetilde{\mathbf{A}}_i \mathbf{W} \, d\Omega + \int_{\Gamma+} \mathbf{W} \cdot \widetilde{\mathbf{A}}_n \mathbf{W} \, d\Gamma \end{aligned} \tag{4.90}$$

In (4.90),

$$\begin{aligned} \mathbf{W}_{,i} \cdot \widetilde{\mathbf{A}}_i \mathbf{W} &= (\mathbf{W} \cdot \widetilde{\mathbf{A}}_i \mathbf{W})_{,i} - \mathbf{W} \cdot (\widetilde{\mathbf{A}}_i \mathbf{W})_{,i} \\ &= (\mathbf{W} \cdot \widetilde{\mathbf{A}}_i \mathbf{W})_{,i} - \mathbf{W} \cdot \widetilde{\mathbf{A}}_{i,i} \mathbf{W} - \mathbf{W} \cdot \widetilde{\mathbf{A}}_i \mathbf{W}_{,i} \\ &= (\mathbf{W} \cdot \widetilde{\mathbf{A}}_i \mathbf{W})_{,i} - \widetilde{\mathbf{A}}_i^T \mathbf{W} \cdot \mathbf{W}_{,i} \\ &= (\mathbf{W} \cdot \widetilde{\mathbf{A}}_i \mathbf{W})_{,i} - \mathbf{W}_{,i} \cdot \widetilde{\mathbf{A}}_i \mathbf{W} \end{aligned} \tag{4.91}$$

Therefore,

$$\mathbf{W}_{,i} \cdot \widetilde{\mathbf{A}}_i \mathbf{W} = \frac{1}{2} (\mathbf{W} \cdot \widetilde{\mathbf{A}}_i \mathbf{W})_{,i} \tag{4.92}$$

and

$$B(\mathbf{W}, \mathbf{W}) = -\frac{1}{2} \int_\Gamma \mathbf{W} \cdot \widetilde{\mathbf{A}}_n \mathbf{W} \, d\Gamma + \int_{\Gamma+} \mathbf{W} \cdot \widetilde{\mathbf{A}}_n \mathbf{W} \, d\Gamma \tag{4.93}$$

$$= \frac{1}{2} \int_{\Gamma+} \mathbf{W} \cdot \widetilde{\mathbf{A}}_n \mathbf{W} \, d\Gamma \geq 0 \tag{4.94}$$

In the scalar case, we got $\frac{1}{2} \int_{\Gamma+} w \mathbf{u} \cdot \mathbf{n} w \, d\Gamma$.

Galerkin and GLS Discretisations

In previous sections, we examined the basic properties of the weak form. We can now define both Galerkin and GLS discretisations:

- Galerkin

$$B(\mathbf{W}^h, \mathbf{V}^h) = L(\mathbf{W}^h) \tag{4.95}$$

- Galerkin/least-squares

$$B_{\mathrm{GLS}}(\mathbf{W}^h, \mathbf{V}^h) = L_{\mathrm{GLS}}(\mathbf{W}^h) \tag{4.96}$$

where

$$B_{\mathrm{GLS}}(\mathbf{W}^h, \mathbf{V}^h) = B(\mathbf{W}^h, \mathbf{V}^h) + \int_\Omega (\tau \widetilde{\mathcal{L}} \mathbf{W}^h) \cdot \widetilde{\mathcal{L}} \mathbf{V}^h \, d\Omega \tag{4.97}$$

$$L_{\mathrm{GLS}}(\mathbf{W}^h) = L(\mathbf{W}^h) + \int_\Omega (\tau \widetilde{\mathcal{L}} \mathbf{W}^h) \cdot \widetilde{\mathcal{F}} \, d\Omega \tag{4.98}$$

where

$$\widetilde{\mathcal{L}} \mathbf{W}^h = \widetilde{\mathbf{A}} \cdot \nabla \mathbf{W}^h - \nabla \cdot \widetilde{\mathbf{K}} \nabla \mathbf{W}^h \tag{4.99}$$

In (4.97) and (4.98), τ is an $m \times m$ matrix, with the properties of symmetry and positive definiteness, but which is typically not diagonal.

The consistency of GLS is obvious. Its stability is given by:

$$|||\mathbf{W}^h|||^2_{\mathrm{GLS}} \equiv B_{\mathrm{GLS}}(\mathbf{W}^h, \mathbf{W}^h) = |||\mathbf{W}^h|||^2 + ||\tau^{1/2} \widetilde{\mathcal{L}} \mathbf{W}^h||^2_{\Omega'} \tag{4.100}$$

Since τ is a symmetric, positive definite matrix, it possesses a unique symmetric positive definite square-root. In the scalar case, we had $|||w^h|||^2 + ||\tau^{1/2} \mathcal{L} w^h||^2_{\Omega'}$. We now claim that τ can be defined such that

$$|||\mathbf{E}^h|||_{\mathrm{GLS}} = \begin{cases} \mathrm{O}(h^k) & \text{in the parabolic case} \\ \mathrm{O}(h^{k+1/2}) & \text{in the hyperbolic case} \end{cases} \tag{4.101}$$

4.2.3 Unsteady Case

In the unsteady case, we keep the same boundary conditions as in the steady case, together with the following additional initial conditions:

$$\mathbf{V}(\mathbf{x}, 0) = \mathbf{V}_0(\mathbf{x}) \quad , \quad \forall \mathbf{x} \in \Omega \tag{4.102}$$

In the parabolic and hyperbolic cases, the discontinuous in time Galerkin formulation is defined by

$$B(\mathbf{W}^h, \mathbf{V}^h)_n = L(\mathbf{W}^h)_n \tag{4.103}$$

where

$$\begin{aligned} B(\mathbf{W}^h, \mathbf{V}^h)_n &= -\int_{t_n}^{t_{n+1}} \int_\Omega \dot{\mathbf{W}}^h \cdot \widetilde{\mathbf{A}}_0 \mathbf{V}^h \, d\Omega \, dt + \int_{t_n}^{t_{n+1}} B(\mathbf{W}^h, \mathbf{V}^h) \, dt \\ &\quad + \left(\mathbf{W}^h(t_{n+1}^-), \widetilde{\mathbf{A}}_0 \mathbf{V}^h(t_{n+1}^-) \right)_\Omega \end{aligned} \tag{4.104}$$

$$L(\mathbf{W}^h)_n = \int_{t_n}^{t_{n+1}} L(\mathbf{W}^h) \, dt + \left(\mathbf{W}^h(t_n^+), \widetilde{\mathbf{A}}_0 \mathbf{V}^h(t_n^-) \right)_\Omega \tag{4.105}$$

On the n^{th} time slab, consistency is assured by

$$\begin{aligned}
B(\mathbf{W}^h, \mathbf{V})_n - L(\mathbf{W}^h)_n &= \int_{t_n}^{t_{n+1}} \int_{\Omega'} \mathbf{W}^h \cdot \left(\widetilde{\mathcal{L}}_t \mathbf{V} - \widetilde{\mathcal{F}} \right) \, d\Omega \, dt \\
&\quad + \int_{\Omega} \mathbf{W}^h(t_n^+) \cdot \left(\widetilde{\mathbf{A}}_0 [\![\mathbf{V}(t_n)]\!] \right) \, d\Omega \\
&= 0
\end{aligned} \tag{4.106}$$

where

$$\widetilde{\mathcal{L}}_t \mathbf{V} = \widetilde{\mathbf{A}}_0 \mathbf{V}_{,t} + \widetilde{\mathcal{L}} \mathbf{V} \tag{4.107}$$

The Galerkin/least-squares counterpart is defined by:

$$B_{\text{GLS}}(\mathbf{W}^h, \mathbf{V}^h)_n = L_{\text{GLS}}(\mathbf{W}^h)_n \tag{4.108}$$

with

$$B_{\text{GLS}}(\mathbf{W}^h, \mathbf{V}^h)_n = B(\mathbf{W}^h, \mathbf{V}^h)_n + \left(\tau \widetilde{\mathcal{L}}_t \mathbf{W}^h, \widetilde{\mathcal{L}}_t \mathbf{V}^h \right)_{\tilde{\mathsf{Q}}_n} \tag{4.109}$$

$$L_{\text{GLS}}(\mathbf{W}^h)_n = L(\mathbf{W}^h)_n + \left(\tau \widetilde{\mathcal{L}}_t \mathbf{W}^h, \widetilde{\mathcal{F}} \right)_{\tilde{\mathsf{Q}}_n} \tag{4.110}$$

For the purpose of the convergence analysis, we globalize the linear and bilinear form defined above:

$$\mathbf{B}(\mathbf{W}^h, \mathbf{V}^h) = \sum_{n=0}^{n_{\max}-1} \left(B(\mathbf{W}^h, \mathbf{V}^h)_n - \left(\mathbf{W}^h(t_n^+), \widetilde{\mathbf{A}}_0 \mathbf{V}^h(t_n^-) \right)_{\Omega} \right) \tag{4.111}$$

$$\mathbf{L}(\mathbf{W}^h) = \sum_{n=0}^{n_{\max}-1} \left(L(\mathbf{W}^h)_n - \left(\mathbf{W}^h(t_n^+), \widetilde{\mathbf{A}}_0 \mathbf{V}^h(t_n^-) \right)_{\Omega} \right) \tag{4.112}$$

$$= \sum_{n=0}^{n_{\max}-1} \int_{t_n}^{t_{n+1}} L(\mathbf{W}^h) \, dt \tag{4.113}$$

Exercise 4.3 (Stability for the Galerkin method)
Show that

$$|\!|\!| \mathbf{W}^h |\!|\!| \equiv \mathbf{B}(\mathbf{W}^h, \mathbf{W}^h) \tag{4.114}$$

where

$$\begin{aligned}
|\!|\!| \mathbf{W}^h |\!|\!| &= \frac{1}{2} \Big(|| \widetilde{\mathbf{A}}_0^{1/2} \mathbf{W}^h(T^-) ||_{\Omega}^2 + || \widetilde{\mathbf{A}}_0^{1/2} \mathbf{W}^h(0^+) ||_{\Omega}^2 \\
&\quad + \sum_{n=1}^{n_{\max}-1} || \widetilde{\mathbf{A}}_0 [\![\mathbf{W}^h(t_n)]\!] ||_{\Omega}^2 \Big) + \int_0^T |||\mathbf{W}^h|||^2 \, dt
\end{aligned}$$

Exercise 4.4 (Consistency for the Galerkin method)
Show that

$$\mathbf{B}(\mathbf{W}^h, \mathbf{V}) = \mathbf{L}(\mathbf{W}^h) \tag{4.115}$$

and conclude that

$$\mathbf{B}(\mathbf{W}^h, \mathbf{E}) = 0 \tag{4.116}$$

In the same fashion, we globalize GLS:

$$\begin{aligned} \mathbf{B}_{\mathrm{GLS}}(\mathbf{W}^h, \mathbf{V}^h) &= \mathbf{B}(\mathbf{W}^h, \mathbf{V}^h) + (\tau \widetilde{\mathcal{L}}_t \mathbf{W}^h, \widetilde{\mathcal{L}}_t \mathbf{V}^h)_{\tilde{\mathsf{Q}}} && (4.117) \\ \mathbf{L}_{\mathrm{GLS}}(\mathbf{W}^h) &= \mathbf{L}(\mathbf{W}^h) + (\tau \widetilde{\mathcal{L}}_t \mathbf{W}^h, \widetilde{\mathcal{F}})_{\tilde{\mathsf{Q}}} && (4.118) \end{aligned}$$

The stability statement for GLS reads:

$$\begin{aligned} |\!|\!| \mathbf{W}^h |\!|\!|^2_{\mathrm{GLS}} &\equiv \mathbf{B}_{\mathrm{GLS}}(\mathbf{W}^h, \mathbf{W}^h) \\ &= |\!|\!| \mathbf{W}^h |\!|\!| + || \tau^{1/2} \widetilde{\mathcal{L}}_t \mathbf{W}^h ||^2_{\tilde{\mathsf{Q}}} && (4.119) \end{aligned}$$

We obviously also have consistency:

$$\mathbf{B}_{\mathrm{GLS}}(\mathbf{W}^h, \mathbf{E}) = 0 \tag{4.120}$$

One can show the following error estimate:

$$|\!|\!| \mathbf{E}^h |\!|\!|_{\mathrm{GLS}} = \begin{cases} \mathrm{O}(h^k) & \text{in the parabolic case} \\ \mathrm{O}(h^{k+1/2}) & \text{in the hyperbolic case} \end{cases} \tag{4.121}$$

A sketch of the convergence analysis for advective-diffusive systems follows:

We have

$$\begin{aligned} \mathbf{E} = \mathbf{V}^h - \mathbf{V} &= \mathbf{V}^h - \tilde{\mathbf{V}}^h + \tilde{\mathbf{V}}^h - \mathbf{V} \\ &= \mathbf{E}^h + \mathbf{H} && (4.122) \end{aligned}$$

Thus

$$\begin{aligned} |\!|\!| \mathbf{E}^h |\!|\!|^2_{\mathrm{GLS}} &= \mathbf{B}_{\mathrm{GLS}}(\mathbf{E}^h, \mathbf{E}^h) \\ &\leq \left| \mathbf{B}_{\mathrm{GLS}}(\mathbf{E}^h, \mathbf{H}) \right| \\ &= \left| \mathbf{B}(\mathbf{E}^h, \mathbf{H}) + (\tau \mathcal{L}_t \mathbf{E}^h, \mathcal{L}_t \mathbf{H})_{\tilde{\mathsf{Q}}} \right| && (4.123) \end{aligned}$$

where

$$
\begin{aligned}
\mathbf{B}(\mathbf{E}^h, \mathbf{H}) &= \sum_{n=0}^{n_{\max}-1} \left\{ B(\mathbf{E}^h, \mathbf{H})_n - (\mathbf{E}^h(t_n^+), \widetilde{\mathbf{A}}_0 \mathbf{H}(t_n^-))_\Omega \right\} \\
&= \sum_{n=0}^{n_{\max}-1} \left\{ \int_{t_n}^{t_{n+1}} \left(-(\dot{\mathbf{E}}^h, \widetilde{\mathbf{A}}_0 \mathbf{H})_\Omega + B(\mathbf{E}^h, \mathbf{H}) \right) \, dt \right. \\
&\quad \left. +(\mathbf{E}^h(t_{n+1}^-), \widetilde{\mathbf{A}}_0 \mathbf{H}(t_{n+1}^-))_\Omega - (\mathbf{E}^h(t_n^+), \widetilde{\mathbf{A}}_0 \mathbf{H}(t_n^-))_\Omega \right\} \\
&= \sum_{n=0}^{n_{\max}-1} \int_{t_n}^{t_{n+1}} \left(-(\dot{\mathbf{E}}^h, \widetilde{\mathbf{A}}_0 \mathbf{H})_\Omega + B(\mathbf{E}^h, \mathbf{H}) \right) \, dt \\
&\quad +(\mathbf{E}^h(T^-), \widetilde{\mathbf{A}}_0 \mathbf{H}(T^-))_\Omega - (\mathbf{E}^h(0^+), \widetilde{\mathbf{A}}_0 \mathbf{H}(0^-))_\Omega \\
&\quad - \sum_{n=1}^{n_{\max}-1} (\llbracket \mathbf{E}^h(t_n) \rrbracket, \widetilde{\mathbf{A}}_0 \mathbf{H}(t_n^-))_\Omega
\end{aligned}
\tag{4.124}
$$

Exercise 4.5 *Verify eq. (4.124).*

- Let us examine the case $\widetilde{\mathbf{K}} > \mathbf{0}$ first. In this case, we get

$$
\begin{aligned}
B(\mathbf{E}^h, \mathbf{H}) &= \int_\Omega \left(-\nabla \mathbf{E}^h \cdot \widetilde{\mathbf{A}} \mathbf{H} + \nabla \mathbf{E}^h \cdot \widetilde{\mathbf{K}} \nabla \mathbf{H} \right) \, d\Omega \\
&= -(\nabla \mathbf{E}^h, \widetilde{\mathbf{A}} \mathbf{H})_\Omega + (\nabla \mathbf{E}^h, \widetilde{\mathbf{K}} \nabla \mathbf{H})_\Omega \\
&= -(\widetilde{\mathbf{A}} \cdot \nabla \mathbf{E}^h, \mathbf{H})_\Omega + (\nabla \mathbf{E}^h, \widetilde{\mathbf{K}} \nabla \mathbf{H})_\Omega
\end{aligned}
\tag{4.125}
$$

Thus,

$$\begin{aligned}
\mathbf{B}(\mathbf{E}^h, \mathbf{H}) &= \sum_{n=0}^{n_{\max}-1} \int_{t_n}^{t_{n+1}} \Big(-(\widetilde{\mathbf{A}}_0 \dot{\mathbf{E}}^h + \widetilde{\mathbf{A}} \cdot \nabla \mathbf{E}^h, \mathbf{H})_\Omega + (\nabla \cdot (\widetilde{\mathbf{K}} \nabla \mathbf{E}^h), \mathbf{H})_{\Omega'} \\
&\quad -(\nabla \cdot (\widetilde{\mathbf{K}} \nabla \mathbf{E}^h), \mathbf{H})_{\Omega'} + (\nabla \mathbf{E}^h, \widetilde{\mathbf{K}} \nabla \mathbf{H})_\Omega \Big) \, dt \\
&\quad +(\mathbf{E}^h(T^-), \widetilde{\mathbf{A}}_0 \mathbf{H}(T^-))_\Omega - (\mathbf{E}^h(0^+), \widetilde{\mathbf{A}}_0 \mathbf{H}(0^-))_\Omega \\
&\quad - \sum_{n=1}^{n_{\max}-1} ([\![\mathbf{E}^h(t_n)]\!], \widetilde{\mathbf{A}}_0 \mathbf{H}(t_n^-))_\Omega \\
&= \sum_{n=0}^{n_{\max}-1} \int_{t_n}^{t_{n+1}} \Big(-(\widetilde{\mathcal{L}}_t \mathbf{E}^h, \mathbf{H})_{\Omega'} - (\nabla \cdot \widetilde{\mathbf{K}} \nabla \mathbf{E}^h, \mathbf{H})_{\Omega'} + (\nabla \mathbf{E}^h, \widetilde{\mathbf{K}} \nabla \mathbf{H})_\Omega \Big) \, dt \\
&\quad +(\mathbf{E}^h(T^-), \widetilde{\mathbf{A}}_0 \mathbf{H}(T^-))_\Omega - (\mathbf{E}^h(0^+), \widetilde{\mathbf{A}}_0 \mathbf{H}(0^-))_\Omega \\
&\quad - \sum_{n=1}^{n_{\max}-1} ([\![\mathbf{E}^h(t_n)]\!], \widetilde{\mathbf{A}}_0 \mathbf{H}(t_n^-))_\Omega \\
&= \sum_{n=0}^{n_{\max}-1} \int_{t_n}^{t_{n+1}} \Big(-(\tau^{1/2} \widetilde{\mathcal{L}}_t \mathbf{E}^h, \tau^{-1/2} \mathbf{H})_{\Omega'} - (\bar{\tau}^{1/2} \nabla \cdot \widetilde{\mathbf{K}} \nabla \mathbf{E}^h, \bar{\tau}^{-1/2} \mathbf{H})_{\Omega'} \\
&\quad +(\widetilde{\mathbf{K}}^{1/2} \nabla \mathbf{E}^h, \widetilde{\mathbf{K}}^{1/2} \nabla \mathbf{H})_\Omega \Big) \, dt \\
&\quad +(\mathbf{E}^h(T^-), \widetilde{\mathbf{A}}_0 \mathbf{H}(T^-))_\Omega - (\mathbf{E}^h(0^+), \widetilde{\mathbf{A}}_0 \mathbf{H}(0^-))_\Omega \\
&\quad - \sum_{n=1}^{n_{\max}-1} ([\![\mathbf{E}^h(t_n)]\!], \widetilde{\mathbf{A}}_0 \mathbf{H}(t_n^-))_\Omega
\end{aligned} \tag{4.126}$$

Therefore, for all $\{\epsilon_i\}_{1 \le i \le 7}$, we get

$$\begin{aligned}
|\!|\!|\mathbf{E}^h|\!|\!|^2_{\mathrm{GLS}} &\le \frac{\epsilon_1}{2} ||\widetilde{\mathbf{A}}_0^{1/2} \mathbf{E}^h(T^-)||^2_\Omega + \frac{1}{2\epsilon_1} ||\widetilde{\mathbf{A}}_0^{1/2} \mathbf{H}(T^-)||^2_\Omega \\
&\quad + \frac{\epsilon_2}{2} ||\widetilde{\mathbf{A}}_0^{1/2} \mathbf{E}^h(0^+)||^2_\Omega + \frac{1}{2\epsilon_2} ||\widetilde{\mathbf{A}}_0^{1/2} \mathbf{H}(0^-)||^2_\Omega \\
&\quad + \sum_{n=1}^{n_{\max}-1} \Big\{ \frac{\epsilon_3}{2} ||\widetilde{\mathbf{A}}_0^{1/2} [\![\mathbf{E}^h(t_n)]\!]||^2_\Omega + \frac{1}{2\epsilon_3} ||\widetilde{\mathbf{A}}_0^{1/2} \mathbf{H}(t_n^-)||^2_\Omega \Big\} \\
&\quad + \frac{\epsilon_4}{2} ||\tau^{1/2} \widetilde{\mathcal{L}}_t \mathbf{E}^h||^2_{\tilde{\mathsf{Q}}} + \frac{1}{2\epsilon_4} ||\tau^{-1/2} \mathbf{H}||^2_{\mathsf{Q}} \\
&\quad + \frac{\epsilon_5}{2} ||\bar{\tau}^{1/2} \nabla \cdot \widetilde{\mathbf{K}} \nabla \mathbf{E}^h||^2_{\tilde{\mathsf{Q}}} + \frac{1}{2\epsilon_5} ||\bar{\tau}^{-1/2} \mathbf{H}||^2_{\mathsf{Q}} \\
&\quad + \frac{\epsilon_6}{2} ||\widetilde{\mathbf{K}}^{1/2} \nabla \mathbf{E}^h||^2_{\mathsf{Q}} + \frac{1}{2\epsilon_6} ||\widetilde{\mathbf{K}}^{1/2} \nabla \mathbf{H}||^2_{\mathsf{Q}} \\
&\quad + \frac{\epsilon_7}{2} ||\tau^{1/2} \widetilde{\mathcal{L}}_t \mathbf{E}^h||^2_{\tilde{\mathsf{Q}}} + \frac{1}{2\epsilon_7} ||\tau^{1/2} \widetilde{\mathcal{L}}_t \mathbf{H}||^2_{\tilde{\mathsf{Q}}}
\end{aligned} \tag{4.127}$$

We need the properties

1. $||\bar{\tau}^{1/2}\nabla \cdot \widetilde{\mathbf{K}}\nabla \mathbf{E}^h||^2_{\tilde{\mathsf{Q}}} \leq ||\widetilde{\mathbf{K}}^{1/2}\mathbf{E}^h||^2_{\mathsf{Q}}$ (inverse estimate)
2. $||\bar{\tau}^{-1/2}\mathbf{H}||^2_{\mathsf{Q}} = O(h^{2k})$.

$\bar{\tau}^{1/2}$ needs to gain back the powers of h lost in the inverse estimate. Therefore, in this case, $\bar{\tau}$ needs to be $O(h^2)$. But, in the advection dominated case, $\widetilde{\mathbf{K}} = O(h)$, and $\bar{\tau} = O(h)$ because we need $||\bar{\tau}^{-1/2}\mathbf{H}||^2_{\mathsf{Q}} = O(h^{2k+1})$. Such τ's have been defined in Hughes and Mallet [130] and Shakib, Hughes and Johan [224].

Remark 4.18 *A mistake made in these references is the failure to note that $\bar{\tau}$ is distinct from τ. An awareness of this comes from the scalar case where τ did not appear in the analogous terms:*

$$\kappa(e^h_{,xx}, \eta) \leq \frac{\kappa}{2}(\epsilon||e^h_{,xx}||^2 + \frac{1}{\epsilon}||\eta||^2) \tag{4.128}$$

The inverse estimate $||e^h_{,xx}||_{\Omega'} \leq c_I h^{-1}||e^h_{,x}||$ led to $\epsilon = \frac{1}{2}c_I^2 h^2$.

In this case, we choose $\epsilon_i = \frac{1}{2}$ for $1 \leq i \leq 7$. Therefore, the $\mathbf{E}^h$-terms on the right-hand side being exactly the half of those of the left-hand side in $|\!|\!|\mathbf{E}^h|\!|\!|^2_{\mathrm{GLS}}$, we get

$$\begin{aligned} \frac{1}{2}|\!|\!|\mathbf{E}^h|\!|\!|^2_{\mathrm{GLS}} \leq \;& ||\widetilde{\mathbf{A}}_0^{1/2}\mathbf{H}(T-)||^2_{\Omega} + ||\widetilde{\mathbf{A}}_0^{1/2}\mathbf{H}(0-)||^2_{\Omega} \\ &+ \sum_{n=1}^{n_{\max}-1} ||\widetilde{\mathbf{A}}_0^{1/2}\mathbf{H}(t_n^-)||^2_{\Omega} + ||\tau^{-1/2}\mathbf{H}||^2_{\mathsf{Q}} \\ &+ ||\tau^{1/2}\widetilde{\mathcal{L}}_t\mathbf{H}||^2_{\tilde{\mathsf{Q}}} + ||\bar{\tau}^{-1/2}\mathbf{H}||^2_{\mathsf{Q}} \\ &+ ||\widetilde{\mathbf{K}}^{1/2}\nabla\mathbf{H}||^2_{\mathsf{Q}} \end{aligned} \tag{4.129}$$

Consequently,

1. In the diffusion dominated case,

$$|\!|\!|\mathbf{E}|\!|\!|_{\mathrm{GLS}} = O(h^k) \tag{4.130}$$

2. In the advection dominated case,

$$|\!|\!|\mathbf{E}|\!|\!|_{\mathrm{GLS}} = O(h^{k+1/2}) \tag{4.131}$$

- In the case where $\widetilde{\mathbf{K}} = \mathbf{0}$, we just have to add the boundary integral term over Γ^+.

$$\begin{aligned} |\!|\!|\mathbf{E}^h|\!|\!|^2_{\mathrm{GLS}} &= \text{as before} + \int_0^T \frac{1}{2}\int_{\Gamma^+} \mathbf{E}^h \cdot \widetilde{\mathbf{A}}_n\mathbf{E}^h \, d\Gamma \, dt \\ &= \mathbf{B}_{\mathrm{GLS}}(\mathbf{E}^h, \mathbf{E}^h) \\ &\leq |\mathbf{B}_{\mathrm{GLS}}(\mathbf{E}^h, \mathbf{H})| \end{aligned} \tag{4.132}$$

This leads to

$$\begin{aligned} \frac{1}{2}\|\mathbf{E}^h\|^2_{\mathrm{GLS}} \leq\ & ||\widetilde{\mathbf{A}}_0^{1/2}\mathbf{H}(T^-)||^2_\Omega + ||\widetilde{\mathbf{A}}_0^{1/2}\mathbf{H}(0^-)||^2_\Omega \\ & + \sum_{n=1}^{n_{\max}-1} ||\widetilde{\mathbf{A}}_0^{1/2}\mathbf{H}(t_n^-)||^2_\Omega + ||\tau^{-1/2}\mathbf{H}||^2_{\mathsf{Q}} \\ & + ||\tau^{1/2}\widetilde{\mathcal{L}}_t\mathbf{H}||^2_{\tilde{\mathsf{Q}}} \\ & + \int_0^T \int_{\Gamma^+} \mathbf{H}\cdot\widetilde{\mathbf{A}}_n\mathbf{H}\, d\Gamma\, dt \end{aligned} \tag{4.133}$$

We get

$$\|\mathbf{E}\|_{\mathrm{GLS}} = O(h^{k+1/2}) \tag{4.134}$$