

# Chapter 4

## Variational and Stabilized Finite Element Methods

### 4.1 Introduction

Before proceeding with the variational formulation of the fluid-structure coupled system, let us look at the convection-diffusion-reaction (CDR) equation which forms a canonical equation for any continuum transport system. The present chapter discusses the variational formulation and finite element technique applied to the CDR equation and reviews various types of stabilization methods.

The CDR equation is given as:

$$\frac{\partial \varphi}{\partial t} + \mathbf{v} \cdot \nabla \varphi - \nabla \cdot (\mathbf{k} \nabla \varphi) + s\varphi = f, \quad (4.1)$$

on a  $d$ -dimensional domain  $\Omega(t) \subset \mathbb{R}^d$ , where  $\varphi$  is the unknown transport variable,  $\mathbf{v}$ ,  $\mathbf{k}$ ,  $s$  and  $f$  are the convection velocity, diffusivity tensor ( $\mathbf{k} = k\mathbf{I}$  for isotropic diffusion,  $k$  and  $\mathbf{I}$  being the diffusion coefficient and identity tensor respectively), reaction coefficient and source respectively.

Numerical discretization of the domain cannot capture the continuum effects properly, thus leading to numerical errors in the solution. The various effects in the CDR equation are (a) convection, which is characterized by the first-order spatial derivatives and transports any information along the characteristic curves with velocity  $\mathbf{v}$ , (b) diffusion, which is represented by the second-order spatial derivatives and spreads the information through the domain with the diffusion coefficient  $\mathbf{k}$ , and (c) reaction, denoted by  $s\varphi$  which is the cause of any production or destruction in the information.

The roots of the characteristic polynomial of Eq. (4.1) in one-dimension (the sign of  $D = |\mathbf{v}|^2 + 4ks$ ) can be utilized to characterize the exact solution. When  $D \geq 0$ , the solution is said to be in the exponential regime, with production for  $s < 0$  and dissipation (or destruction) for  $s \geq 0$ . It constitutes the propagation regime when  $D < 0$ . The solution exhibits distinctive properties based on the dominant effects of the CDR equation and the boundary conditions. A boundary layer-like behavior may be observed near the regions where high gradients of the solution exist. At the

outflow boundaries, exponential layers are formed, while parabolic or characteristic boundary layers are formed along the tangent to the flow direction. Any sharp discontinuity inside the computational domain can also induce such layers, called parabolic or characteristic internal layers [202].

Numerical solution with the standard Galerkin weighted residual method leads to spurious oscillations when the effects are convection dominant [93]. These oscillations are due to the inability of the discretization to resolve the high gradients. To stabilize these oscillations, several methods have been proposed in the literature called Petrov-Galerkin stabilization where the weighting function is perturbed based on a stabilization parameter  $\tau$ , which tends to capture the missing subgrid scale physics. The perturbation acts as an upwinding function and removes the effect of the outflow boundary condition on the convection term. Some of the widely used techniques are streamline upwind Petrov-Galerkin (SUPG) [34], Galerkin/least-squares (GLS) [97] and subgrid scale (SGS) [48] methods. A detailed discussion about the development of these types of stabilization techniques in the literature can be found in Appendix A4.2.

In the upcoming sections, we discuss widely used linear stabilization methods (SUPG, GLS and SGS) for the discretization of the CDR equation and their numerical properties. We then derive the nonlinear expressions to impart the positivity condition at the element matrix level to form the positivity preserving variational (PPV) method [112] and generalize it to different types of mesh discretizations.

## 4.2 The Convection-Diffusion-Reaction Equation

The CDR equation forms a canonical form for complex transport systems like Navier-Stokes, turbulence, two-phase equations and so on. Here, we first review the semi-discrete variational form of the equation.

### 4.2.1 Strong Differential Form

Consider a  $d$ -dimensional spatial domain  $\Omega(t) \subset \mathbb{R}^d$  with the Dirichlet and Neumann boundaries denoted by  $\Gamma_D^\varphi(t)$  and  $\Gamma_N^\varphi(t)$  respectively. The strong form of the CDR equation (with  $\varphi$  as the scalar variable) along with the boundary conditions can be written as

$$\frac{\partial \varphi}{\partial t} + \mathbf{v} \cdot \nabla \varphi - \nabla \cdot (\mathbf{k} \nabla \varphi) + s\varphi = f, \quad \text{on } \Omega(t) \times [0, T], \quad (4.2)$$

$$\varphi = \varphi_D, \quad \text{on } \Gamma_D^\varphi(t) \times [0, T], \quad (4.3)$$

$$\mathbf{k} \nabla \varphi \cdot \mathbf{n}^\varphi = \varphi_N, \quad \text{on } \Gamma_N^\varphi(t) \times [0, T], \quad (4.4)$$

$$\varphi = \varphi_0, \quad \text{on } \Omega(0), \quad (4.5)$$

where  $\varphi = \varphi(\mathbf{x}, t)$  is the scalar unknown which depends on the spatial and temporal coordinates  $\mathbf{x}$  and  $t$  respectively,  $\mathbf{v}$ ,  $\mathbf{k}$ ,  $s$  and  $f$  are the convection velocity, the diffusivity tensor, the reaction coefficient and the source term respectively. Here, we assume that the diffusivity tensor  $\mathbf{k}$  is isotropic, i.e.,  $\mathbf{k} = k\mathbf{I}$ ,  $\mathbf{I}$  being the identity tensor and  $k$  is a scalar positive quantity defined as the diffusion coefficient of the problem. The Dirichlet and Neumann boundary conditions are given by  $\varphi_D$  and  $\varphi_N$  respectively and  $\mathbf{n}^\varphi$  represents the unit normal to the Neumann boundary.  $\varphi_0$  denotes the initial condition on  $\varphi$ .

#### 4.2.2 Semi-Discrete Variational Form

We utilize the generalized- $\alpha$  method [44] to discretize the equation in time. This technique allows user-defined high-frequency damping via controlling a parameter called the spectral radius  $\rho_\infty$ , which is helpful for coarser discretization in space and time. The following expressions are employed for the temporal discretization:

$$\varphi^{n+1} = \varphi^n + \Delta t \partial_t \varphi^n + \gamma \Delta t (\partial_t \varphi^{n+1} - \partial_t \varphi^n), \quad (4.6)$$

$$\partial_t \varphi^{n+\alpha_m} = \partial_t \varphi^n + \alpha_m (\partial_t \varphi^{n+1} - \partial_t \varphi^n), \quad (4.7)$$

$$\varphi^{n+\alpha} = \varphi^n + \alpha (\varphi^{n+1} - \varphi^n), \quad (4.8)$$

where  $\gamma$ ,  $\alpha$  and  $\alpha_m$  are the generalized- $\alpha$  parameters given by

$$\alpha_m = \frac{1}{2} \left( \frac{3 - \rho_\infty}{1 + \rho_\infty} \right), \quad \alpha = \frac{1}{1 + \rho_\infty}, \quad \gamma = \frac{1}{2} + \alpha_m - \alpha. \quad (4.9)$$

In the present chapter, we select the parameters  $\alpha = \alpha_m = \gamma = 0.5$  which correspond to  $\rho_\infty = 1$  so that  $\partial_t \varphi^{n+\alpha_m} = (\varphi^{n+\alpha} - \varphi^n)/(\alpha \Delta t)$ . The temporally discretized CDR equation can thus be written as

$$\partial_t \varphi^{n+\alpha_m} + \mathbf{v} \cdot \nabla \varphi^{n+\alpha} - \nabla \cdot (\mathbf{k} \nabla \varphi^{n+\alpha}) + s \varphi^{n+\alpha} = f(t^{n+\alpha}), \quad (4.10)$$

$$\frac{\varphi^{n+\alpha} - \varphi^n}{\alpha \Delta t} + \mathbf{v} \cdot \nabla \varphi^{n+\alpha} - \nabla \cdot (\mathbf{k} \nabla \varphi^{n+\alpha}) + s \varphi^{n+\alpha} = f(t^{n+\alpha}). \quad (4.11)$$

We can take the temporal derivative coefficient inside the reaction term so that Eq. (4.11) can also be expressed as

$$\mathbf{v} \cdot \nabla \varphi^{n+\alpha} - \nabla \cdot (\mathbf{k} \nabla \varphi^{n+\alpha}) + \left( s + \frac{1}{\alpha \Delta t} \right) \varphi^{n+\alpha} = \left( f + \frac{1}{\alpha \Delta t} \varphi^n \right). \quad (4.12)$$

The above equation can be observed as a steady-state equation with modified reaction coefficient and source term. Let the modified coefficients be given by  $\tilde{\mathbf{v}}$ ,  $\tilde{\mathbf{k}}$ ,  $\tilde{s}$  and  $\tilde{f}$  defined as

$$\tilde{\mathbf{v}} = \mathbf{v}, \quad \tilde{\mathbf{k}} = \mathbf{k}, \quad \tilde{s} = s + \frac{1}{\alpha \Delta t}, \quad \tilde{f} = f + \frac{1}{\alpha \Delta t} \varphi^n. \quad (4.13)$$

Therefore, we will now discretize the following equation in the spatial domain:

$$\tilde{\mathbf{v}} \cdot \nabla \varphi^{n+\alpha} - \nabla \cdot (\tilde{\mathbf{k}} \nabla \varphi^{n+\alpha}) + \tilde{s} \varphi^{n+\alpha} = \tilde{f} \quad \text{on } \Omega(t). \quad (4.14)$$

The domain  $\Omega(t)$  is discretized into  $n_{\text{el}}$  number of elements such that  $\Omega(t) = \bigcup_{e=1}^{n_{\text{el}}} \Omega^e$  and  $\emptyset = \cap_{e=1}^{n_{\text{el}}} \Omega^e$ . The space of trial solution and test function,  $\mathcal{S}_\varphi^h$  and  $\mathcal{V}_\varphi^h$  respectively for the variational formulation are defined as

$$\mathcal{S}_\varphi^h = \{ \varphi_h \mid \varphi_h \in H^1(\Omega(t)), \varphi_h = \varphi_D \text{ on } \Gamma_D^\varphi(t) \}, \quad (4.15)$$

$$\mathcal{V}_\varphi^h = \{ w_h \mid w_h \in H^1(\Omega(t)), w_h = 0 \text{ on } \Gamma_D^\varphi(t) \}. \quad (4.16)$$

These function spaces form the mathematical preliminaries to the variational formulation and their properties have been discussed to some extent in Appendix A4.1. The variational statement for the discretized CDR equation is thus given as: find  $\varphi_h(\mathbf{x}, t^{n+\alpha}) \in \mathcal{S}_\varphi^h$  such that  $\forall w_h \in \mathcal{V}_\varphi^h$ ,

$$\int_{\Omega(t)} \left( w_h (\tilde{\mathbf{v}} \cdot \nabla \varphi_h) - w_h \nabla \cdot (\tilde{\mathbf{k}} \nabla \varphi_h) + w_h \tilde{s} \varphi_h \right) d\Omega = \int_{\Omega(t)} w_h \tilde{f} d\Omega, \quad (4.17)$$

Using the divergence theorem and the fact that  $w_h = 0$  on  $\Gamma_D^\varphi(t)$ , Eq. (4.17) becomes

$$\int_{\Omega(t)} \left( w_h (\tilde{\mathbf{v}} \cdot \nabla \varphi_h) + \nabla w_h \cdot (\tilde{\mathbf{k}} \nabla \varphi_h) + w_h \tilde{s} \varphi_h \right) d\Omega = \int_{\Omega(t)} w_h \tilde{f} d\Omega + \int_{\Gamma_N^\varphi} w_h \varphi_N d\Gamma. \quad (4.18)$$

As a result of spurious global oscillations and instability for convection- and reaction-dominated regimes in the Galerkin finite element method, various stabilization techniques have been proposed in the literature, the most widely used of which are SUPG and GLS methods. The stability is introduced through perturbing the test or weighting function so that the effect of upwinding is achieved. The standard variational formulation for such methods is: find  $\varphi_h(\mathbf{x}, t^{n+\alpha}) \in \mathcal{S}_\varphi^h$  such that  $\forall w_h \in \mathcal{V}_\varphi^h$ :

$$\begin{aligned} & \int_{\Omega(t)} \left( w_h (\tilde{\mathbf{v}} \cdot \nabla \varphi_h) + \nabla w_h \cdot (\tilde{\mathbf{k}} \nabla \varphi_h) + w_h \tilde{s} \varphi_h \right) d\Omega \\ & + \sum_{e=1}^{n_{\text{el}}} \int_{\Omega^e} \mathcal{L}^m w_h \tau (\mathcal{L} \varphi_h - \tilde{f}) d\Omega = \int_{\Omega(t)} w_h \tilde{f} d\Omega + \int_{\Gamma_N^\varphi} w_h \varphi_N d\Gamma, \end{aligned} \quad (4.19)$$

where  $\mathcal{L}^m$  is the operator on the weighting function given in Table 4.1 and the expression for the stabilization parameter  $\tau$  is [193]

$$\tau = \left[ \left( \frac{1}{\alpha \Delta t} \right)^2 + 9 \left( \frac{4\tilde{k}}{h^2} \right)^2 + \left( \frac{2|\tilde{\mathbf{v}}|}{h} \right)^2 + \tilde{s}^2 \right]^{-1/2}, \quad (4.20)$$

where  $h$  is the characteristic element length and  $|\tilde{\mathbf{v}}|$  is the magnitude of the convection velocity. The formula for  $\tau$  has been extensively studied in the literature with several variations, and can be estimated through error analysis. The generality of the expression is a topic of discussion in Section 4.3.2.4. The residual of the CDR equation is defined as

$$\mathcal{R}(\varphi_h) = \tilde{\mathbf{v}} \cdot \nabla \varphi_h - \nabla \cdot (\tilde{\mathbf{k}} \nabla \varphi_h) + \tilde{s} \varphi_h - \tilde{f} = \tilde{\mathcal{L}} \varphi_h - \tilde{f}, \quad (4.21)$$

where  $\tilde{\mathcal{L}}$  is the differential operator corresponding to the differential Eq. (4.14).

**Table 4.1** Differential operators on the weighting function for stabilization methods.

Method	Stabilization operator ( $\mathcal{L}^m$ )
SUPG	$\mathcal{L}_{adv} = \tilde{\mathbf{v}} \cdot \nabla$
GLS	$\tilde{\mathcal{L}} = \tilde{\mathbf{v}} \cdot \nabla - \nabla \cdot (\tilde{\mathbf{k}} \nabla) + \tilde{s}$
SGS	$-\tilde{\mathcal{L}}^* = \tilde{\mathbf{v}} \cdot \nabla + \nabla \cdot (\tilde{\mathbf{k}} \nabla) - \tilde{s}$

## 4.3 The Positivity Preserving Variational (PPV) Method

We discuss how we can impart the positivity property to the variational formulation in this section. We design the nonlinear stabilization terms to reduce the oscillations which remain even after the linear stabilization of SUPG or GLS discussed in the previous section.

### 4.3.1 Linear Stabilization

We begin by analyzing the linear stabilization methods (SUPG, GLS and SGS) with respect to the effect of the sign of the reaction coefficient ( $\tilde{s}$ ), owing to the destruction or production effects. Fourier analysis of the discretized methods (GLS and SGS) [112] showed that the SGS method performs well when  $\tilde{s}$  is negative, but, it loses accuracy when  $\tilde{s} >> 0$  due to excessive dissipation [87]. On the other hand, the GLS method is not as diffusive as SGS when  $\tilde{s} \geq 0$ , but it suffers from phase error when  $\tilde{s} < 0$ . Thus, we select a combination of these methods, so that the formulation is benefited in both production and destruction regimes. Note that the effect of the diffusion term is assumed negligible in the differential operator owing to the use of linear and multilinear finite elements. The second term in Eq. (4.19) can thus be

modified as

$$\sum_{e=1}^{n_{el}} \int_{\Omega^e} \mathcal{L} w_h \tau (\tilde{\mathcal{L}} \varphi_h - \tilde{f}) d\Omega, \quad (4.22)$$

where  $\mathcal{L} = \tilde{\mathbf{v}} \cdot \nabla - \nabla \cdot (\tilde{\mathbf{k}} \nabla) + |\tilde{s}|$ . Here, the absolute value function  $\mathcal{F}(x) = |x|$  is defined as

$$\mathcal{F}(x) = \begin{cases} x, & \text{if } x > 0, \\ 0, & \text{if } x = 0, \\ -x, & \text{if } x < 0. \end{cases} \quad (4.23)$$

As can be easily observed, the above method will behave as GLS for  $\tilde{s} \geq 0$  and as SGS when  $\tilde{s} < 0$ . We next design the nonlinear stabilization terms to maintain the positivity property in the variational procedure.

### 4.3.2 Positivity and Nonlinear Stabilization

As mentioned earlier, SUPG and GLS methods are successful in the reduction of the spurious oscillations which are observed in the Galerkin formulation, they still exist some overshoots and undershoots in the solution near the regions of high gradients or a sharp discontinuity in the solution. This is a result of the fact that neither of two methods is monotone or positivity preserving [45]. Physical solutions in nature should preclude such oscillations, maintaining the sign of their neighbor. Therefore, we look into some of the key criteria to enforce the positivity property at the discrete level.

#### 4.3.2.1 The Positivity Condition

To describe the positivity condition, we consider the Galerkin formulation for the convection-dominated problems. The positivity preserving condition for the Eq. (4.2) can be defined by considering a simplified form of the CDR equation with only convection effects:

$$\partial_t \varphi + \mathbf{v} \cdot \nabla \varphi = 0. \quad (4.24)$$

The finite element approximation of an explicit scheme for a one-dimensional element between  $i-1$  and  $i$  nodes can be written as

$$\frac{\varphi_i^{n+1} - \varphi_i^n}{\Delta t} = - \int_{\Omega(t)} w_h \left( u \frac{\partial \varphi_h}{\partial x} \right) d\Omega = - \int_{\Omega(t)} N^T u \frac{\partial N}{\partial x} d\Omega \begin{bmatrix} \varphi_{i-1}^n \\ \varphi_i^n \end{bmatrix}$$

$$= -\frac{u}{2h} \begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} \varphi_{i-1}^n \\ \varphi_i^n \end{bmatrix}, \quad (4.25)$$

where  $N$  is the row-vector of shape functions based on linear Lagrange polynomials for one-dimensional elements satisfying the partition of unity property. Therefore, after assembly of the elements for a uniform grid, the finite element based stencil can be written as follows

$$\frac{\varphi_i^{n+1} - \varphi_i^n}{\Delta t} = -\frac{u}{2h} (\varphi_{i+1}^n - \varphi_{i-1}^n), \quad (4.26)$$

which has a similar structure as that of the central difference scheme. Any scheme that can be written in the form

$$\frac{\varphi_i^{n+1} - \varphi_i^n}{\Delta t} = C^+(\varphi_{i+1}^n - \varphi_i^n) - C^-(\varphi_i^n - \varphi_{i-1}^n), \quad (4.27)$$

satisfies the positivity preserving property if the coefficients  $C^+$  and  $C^-$  satisfy [85]

$$C^+ \geq 0, \quad C^- \geq 0, \quad C^+ + C^- \leq 1. \quad (4.28)$$

Note that the Galerkin approximation in Eq. (4.26) does not satisfy this condition. To impart the positivity property, the scheme has to be modified by addition of the stabilization terms as suggested by [85] for finite difference approximations.

We observed the case with an explicit scheme. Now, let us see what happens in the case of an implicit scheme. An implicit scheme can be expressed as a system of linear equations of the form  $\mathbf{A}\varphi = \mathbf{b}$ , where  $\mathbf{A}$  and  $\mathbf{b}$  are the left-hand side matrix and the right-hand side vector or force vector respectively, with  $\varphi$  being the vector of unknowns. The positivity preservation property can be generalized to the implicit matrix form of the scheme by transforming the matrix  $\mathbf{A} = \{a_{ij}\}$  to an M-matrix which ensures the positivity and convergence. The elements of the M-matrix have the following properties [130]

$$a_{ii} > 0, \forall i, \quad (4.29)$$

$$a_{ij} \leq 0, \forall j \neq i, \quad (4.30)$$

$$\sum_j a_{ij} = 0, \forall i. \quad (4.31)$$

These make up the sufficient conditions for imposing the positivity property to the matrix  $\mathbf{A}$ . The transformation of  $\mathbf{A}$  into an M-matrix is carried out by the addition of the discrete upwind matrix  $\mathbf{D} = \{d_{ij}\}$  to  $\mathbf{A}$ , the elements of which satisfy [131]:

$$d_{ij} = d_{ji} = -\max\{0, a_{ij}, a_{ji}\}, \quad (4.32)$$

$$d_{ii} = -\sum_{j \neq i} d_{ij}. \quad (4.33)$$

*Remark 4.1.* Galerkin as well as linear stabilization methods such as SUPG and GLS are not positivity preserving methods [45]. Addition of the linear stabilization term in SUPG and GLS does not satisfy the required conditions in Eqs. (4.29-4.31) and therefore they are not positivity preserving. The PPV method takes the combination of the GLS-SGS approach in the linear stabilization (which is not positivity preserving) and further adds a nonlinear stabilization term, which is designed in the next section, to impose the positivity property.

#### 4.3.2.2 Enforcement of Positivity Condition in One-Dimension

Let us design the nonlinear stabilization term to impart the positivity property in the linear stabilization method based on combined GLS-SGS methodology in Section 4.3.1. We begin by considering the steady-state CDR equation in one-dimension:

$$u \frac{d\varphi}{dx} - k \frac{d^2\varphi}{dx^2} + s\varphi = f, \quad u \geq 0. \quad (4.34)$$

Assume that only Dirichlet boundary conditions are specified and  $f = 0$  for simplicity. The variational formulation gives: find  $\varphi_h(\mathbf{x}) \in \mathcal{S}_\varphi^h$  such that  $\forall w_h \in \mathcal{V}_\varphi^h$ :

$$\int_{\Omega} \left( w_h u \frac{d\varphi_h}{dx} + \frac{dw_h}{dx} k \frac{d\varphi_h}{dx} + w_h s \varphi_h \right) d\Omega + \sum_{e=1}^{n_{el}} \int_{\Omega^e} \mathcal{L} w_h \tau(\mathcal{L} \varphi_h) d\Omega = 0, \quad (4.35)$$

where  $\mathcal{L}$  is the differential operator of the PPV method, i.e., in one-dimension,  $\mathcal{L} = u \frac{d}{dx} + |s|$  (Here, the higher-order diffusion term is neglected due to the linear finite elements), which gives

$$\begin{aligned} & \int_{\Omega} \left( w_h u \frac{d\varphi_h}{dx} + \frac{dw_h}{dx} k \frac{d\varphi_h}{dx} + w_h s \varphi_h \right) d\Omega \\ & + \sum_{e=1}^{n_{el}} \int_{\Omega^e} \left( u \frac{dw_h}{dx} + |s| w_h \right) \tau \left( u \frac{d\varphi_h}{dx} + s \varphi_h \right) d\Omega = 0. \end{aligned} \quad (4.36)$$

Let  $\mathbf{A}_c^e = [a_{c\{ij\}}^e]$ ,  $\mathbf{A}_d^e = [a_{d\{ij\}}^e]$  and  $\mathbf{A}_r^e = [a_{r\{ij\}}^e]$  be the local element matrices for the first three terms in Eq. (4.36) corresponding to the convection, diffusion and reaction effects respectively. Discretizing the one-dimensional domain using two-node linear elements and  $w_i = N_i$  being the shape functions at the corresponding nodes of the finite element, the matrices can be expressed as

$$a_{c\{ij\}}^e = \int_{\Omega^e} N_i u \frac{dN_j}{dx} d\Omega, \quad a_{d\{ij\}}^e = \int_{\Omega^e} \frac{dN_i}{dx} k \frac{dN_j}{dx} d\Omega, \quad a_{r\{ij\}}^e = \int_{\Omega^e} N_i s N_j d\Omega,$$

which are computed to be

$$\mathbf{A}_c^e = \frac{u}{2} \begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix}, \quad \mathbf{A}_d^e = \frac{k}{h} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}, \quad \mathbf{A}_r^e = \frac{sh}{6} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \quad (4.37)$$

where  $h$  is the characteristic length of the element. Similarly, each expression in Eq. (4.36) can be expressed in the matrix form. The combined element level matrix  $\mathbf{A}^e$  is thus given by

$$\mathbf{A}^e = \begin{bmatrix} -\frac{u}{2} + \frac{k}{h} + \frac{sh}{3} + \frac{\tau u^2}{h} - \frac{u\tau s}{2} - \frac{u\tau|s|}{2} + \frac{\tau s|s|h}{3} & \frac{u}{2} - \frac{k}{h} + \frac{sh}{6} - \frac{\tau u^2}{h} - \frac{u\tau s}{2} + \frac{u\tau|s|}{2} + \frac{\tau s|s|h}{6} \\ -\frac{u}{2} - \frac{k}{h} + \frac{sh}{6} - \frac{\tau u^2}{h} + \frac{u\tau s}{2} - \frac{u\tau|s|}{2} + \frac{\tau s|s|h}{6} & \frac{u}{2} + \frac{k}{h} + \frac{sh}{3} + \frac{\tau u^2}{h} + \frac{u\tau s}{2} + \frac{u\tau|s|}{2} + \frac{\tau s|s|h}{3} \end{bmatrix}. \quad (4.38)$$

The discrete upwind matrix corresponding to  $\mathbf{A}^e$  can be written as

$$\mathbf{D}^e = \begin{bmatrix} \max\left\{\frac{|u-u\tau s+u\tau|s|}{2} - \frac{k+\tau u^2}{h} + \frac{(s+\tau s|s|)h}{6}, 0\right\} & -\max\left\{\frac{|u-u\tau s+u\tau|s|}{2} - \frac{k+\tau u^2}{h} + \frac{(s+\tau s|s|)h}{6}, 0\right\} \\ -\max\left\{\frac{|u-u\tau s+u\tau|s|}{2} - \frac{k+\tau u^2}{h} + \frac{(s+\tau s|s|)h}{6}, 0\right\} & \max\left\{\frac{|u-u\tau s+u\tau|s|}{2} - \frac{k+\tau u^2}{h} + \frac{(s+\tau s|s|)h}{6}, 0\right\} \end{bmatrix} \\ = \frac{k^{\text{add}}}{h} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}, \quad (4.39)$$

where  $k^{\text{add}}$  is

$$k^{\text{add}} = \max\left\{\frac{|u-\tau us+\tau u|s||h}{2} - (k+\tau u^2) + \frac{(s+\tau s|s|)h^2}{6}, 0\right\} \quad (4.40)$$

The discrete upwind matrix  $\mathbf{D}^e$  converts  $\mathbf{A}^e$  to an M-matrix when added to it, thus imparting the positivity property. It can be observed as addition of diffusion to the linear stabilized form. This procedure reduces the order of accuracy of the method to first-order. Therefore, to counteract the loss of accuracy, a nonlinear regulation of this added diffusion is required. This regulation is provided by a nonlinear residual-based solution-dependent parameter which adds diffusion only to the elements where a non-regular solution or oscillations are present. As a result, the loss of accuracy is avoided in the regions where smooth solution is present as the nonlinear term has a small contribution owing to the negligible residual.

The discrete PPV formulation for steady-state CDR equation in one-dimension is: find  $\varphi_h(x) \in \mathcal{S}_\varphi^h$  such that  $\forall w_h \in \mathcal{V}_\varphi^h$ :

$$\int_\Omega \left( w_h \left( u \frac{d\varphi_h}{dx} \right) + k \frac{dw_h}{dx} \frac{d\varphi_h}{dx} + w_h s \varphi_h \right) d\Omega + \sum_{e=1}^{n_{\text{el}}} \int_{\Omega^e} \mathcal{L} w_h \tau (\tilde{\mathcal{L}} \varphi_h - f) d\Omega \\ + \sum_{e=1}^{n_{\text{el}}} \int_{\Omega^e} \chi \frac{|\mathcal{R}(\varphi_h)|}{|\nabla \varphi_h|} k^{\text{add}} \frac{dw_h}{dx} \frac{d\varphi_h}{dx} d\Omega = \int_\Omega w_h f d\Omega + \int_{\Gamma_N^\varphi} w_h \varphi_N d\Gamma, \quad (4.41)$$

where  $\mathcal{R}(\varphi_h)$  is the residual of the given equation and  $\chi$  is a scaling parameter defined using the problem constants ( $u$ ,  $k$ ,  $s$  and  $h$ ) which non-dimensionalizes  $\frac{|\mathcal{R}(\varphi_h)|}{|\nabla \varphi_h|}$ . The added term has been the standard form to add discontinuity capturing term to linear stabilized equation in the literature. Our next step is to derive an expression for  $\chi$ . Let us look into the variation of  $\frac{|\mathcal{R}(\varphi_h)|}{|\nabla \varphi_h|}$  with  $s$ ,  $u$ ,  $k$  and  $h$  inside a finite element. To accomplish this, we consider a pure reaction problem ( $s\varphi = f$ ) and follow

a similar procedure as that in [155]. We obtain a lumped mass matrix after adding the discrete upwind matrix to the left-hand side of the element-level matrix. For such problems, it is known that Gibbs oscillations can be prevented by the lumped mass technique. Therefore, we evaluate  $\frac{|\mathcal{R}(\varphi_h)|}{|\nabla \varphi_h|}$  assuming that the solution tends to that obtained by lumped mass approach. For this purpose, consider a step function as the source  $f(x)$

$$f(x) = \begin{cases} 0, & \forall x \in [0, 0.25 + \eta h], \\ q, & \text{elsewhere,} \end{cases} \quad (4.42)$$

where  $h$  is the element length,  $\eta$  is the location of the discontinuity in the function  $f(x)$ . The lumped mass matrix solution can then be written as

$$\varphi = \left( \frac{q}{s} \right) \left\{ 0, \dots, 0, \frac{(1-\eta)^2}{2}, \frac{(2-\eta^2)}{2}, 1, \dots, 1 \right\}. \quad (4.43)$$

Let us observe how  $\frac{|\mathcal{R}(\varphi_h)|}{|\nabla \varphi_h|}$  behaves inside the element containing the discontinuity at  $\eta$ . For this particular element,

$$\mathcal{R}(\varphi_h) = s\varphi_h - f = q \left( \frac{(1-\xi)(1-\eta)^2}{2} + \frac{(1+\xi)(2-\eta^2)}{2} \right) - f, \quad (4.44)$$

$$\nabla \varphi_h = \frac{q}{2sh}(2\eta - 2\eta^2 + 1), \quad (4.45)$$

which gives

$$\frac{|\mathcal{R}(\varphi_h)|}{|\nabla \varphi_h|} = \left| \frac{\mathcal{R}(\varphi_h)}{\nabla \varphi_h} \right| = \left| \frac{sh}{2} \frac{(1-\xi)(1-\eta)^2 + (1+\xi)(2-\eta^2)}{2\eta - 2\eta^2 + 1} - \frac{2fsh}{q(2\eta - 2\eta^2 + 1)} \right|. \quad (4.46)$$

Generalizing for all the elements,

$$\frac{|\mathcal{R}(\varphi_h)|}{|\nabla \varphi_h|} = \frac{|s\varphi_h - f|}{|\nabla \varphi_h|} = \left| \frac{sh}{2} p(\eta, \xi) \right|, \quad (4.47)$$

where  $p(\eta, \xi)$  is a function depending on the location of the discontinuity  $\eta$  inside an element and the natural coordinate  $\xi$  and is given as

$$p(\eta, \xi) = \begin{cases} \frac{(1-\xi)(1-\eta)^2 + (1+\xi)(2-\eta^2)}{2\eta - 2\eta^2 + 1} - \frac{4f}{q(2\eta - 2\eta^2 + 1)}, & \Omega^e \text{ with discontinuity,} \\ (1+\xi) - \frac{4f}{q(1-\eta)^2}, & \Omega^e \text{ adjoining discontinuity,} \\ 0, & \text{elsewhere.} \end{cases} \quad (4.48)$$

Thus, the nonlinear stabilization term can be expressed as

$$\int_{\Omega^e} \chi \frac{|\mathcal{R}(\phi_h)|}{|\nabla \phi_h|} k^{\text{add}} \frac{dw_h}{dx} \frac{d\phi_h}{dx} d\Omega = \int_{\Omega^e} \chi \left| \frac{sh}{2} p(\eta, \xi) \right| k^{\text{add}} \frac{dw_h}{dx} \frac{d\phi_h}{dx} d\Omega. \quad (4.49)$$

Let the local element matrix after the discretization of the above term be denoted by  $\mathbf{K}_{\text{add}}^e$ . It can be written as

$$\mathbf{K}_{\text{add}}^e = \frac{k^{\text{add}} \chi}{h} \int_{\Omega^e} \left| \frac{sh}{2} p(\eta, \xi) \right| d\Omega \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}. \quad (4.50)$$

After some algebraic manipulations, the above integral can be integrated using the property of the absolute function (Eq. (4.23)) as

$$\int_{\Omega^e} \left| \frac{sh}{2} p(\eta, \xi) \right| d\Omega = \begin{cases} \frac{|s|h^2}{2} r(\eta), & \Omega^e \text{ with discontinuity}, \\ \frac{|s|h^2}{2}, & \Omega^e \text{ adjoining discontinuity}, \\ 0, & \text{elsewhere}, \end{cases} \quad (4.51)$$

where  $r(\eta)$  is a function of the location of the discontinuity inside the element ( $0 \leq \eta \leq 1$ ) given as [155]

$$r(\eta) = \left[ \frac{1 + 2\eta(1 - \eta)[1 - 2\eta(1 - \eta)]}{1 + 2\eta(1 - \eta)} \right]. \quad (4.52)$$

After the evaluation of the ratio  $\frac{|\mathcal{R}(\phi_h)|}{|\nabla \phi_h|}$ , we next look into defining the parameter  $\chi$ . As the change of variable to the local coordinates in the integration introduces Jacobian of the element ( $h/2$  in one-dimensional case) which is geometry-dependent, we need to re-scale  $\chi$  such that it becomes geometry-independent. This leads to scaling the integral as  $\chi = 2/(|s|h)$  so that the regulatory term depends purely on the function  $r(\eta)$ , which is geometry-independent. To extend this concept to various regimes, consider the following cases:

**Case 1:** For the convection-diffusion problem with  $f = 0$ ,

$$\frac{|\mathcal{R}(\phi_h)|}{|\nabla \phi_h|} = \frac{|u \nabla \phi_h|}{|\nabla \phi_h|} = |u|. \quad (4.53)$$

Therefore, we define  $\chi = 1/|u|$ .

**Case 2:** In the diffusion-reaction equation, we construct the same expression for the integral as Eq. (4.49) which gives  $\chi = 2/(|s|h)$ . For linear finite elements, the Laplacian is exactly zero. But, for multilinear elements it is negligible. Therefore, we conclude that for the reaction-dominated problems ( $u \rightarrow 0$ ),  $\chi \rightarrow 2/(|s|h)$  and for the convection-dominated problems ( $s \rightarrow 0$ ),  $\chi \rightarrow 1/|u|$ . Thus,  $\chi$  can be defined as a combined expression as follows:

$$\chi = \frac{2}{|s|h + 2|u|} \quad (4.54)$$

*Remark 4.2.* Note that we have not made any assumption on the sign of the reaction coefficient  $s$ . The definition of the absolute function takes care of the change in the sign of the reaction coefficient in the evaluated integral.

The above derivation completes the design of the nonlinear PPV technique where we impart the positivity condition to the element-level matrix in one-dimension. The bounds for the positivity condition (Eq. (4.28)) for the PPV formulation have been shown for some cases in [112]. We next intend to generalize the procedure to multi-dimensional CDR equation.

#### 4.3.2.3 Extension to Multi-Dimensions

The discrete form of the PPV formulation for the steady-state CDR equation in multi-dimensions can be written as: find  $\varphi_h(\mathbf{x}) \in \mathcal{S}_\varphi^h$  such that  $\forall w_h \in \mathcal{V}_\varphi^h$ :

$$\begin{aligned} & \int_{\Omega} \left( w_h (\mathbf{v} \cdot \nabla \varphi_h) + \nabla w_h \cdot (\mathbf{k} \nabla \varphi_h) + w_h s \varphi_h \right) d\Omega + \sum_{e=1}^{n_{el}} \int_{\Omega_e} \mathcal{L} w_h \tau (\mathcal{L} \varphi_h - f) d\Omega \\ & + \sum_{e=1}^{n_{el}} \int_{\Omega_e} \chi \frac{|\mathcal{R}(\varphi_h)|}{|\nabla \varphi_h|} k_s^{\text{add}} \nabla w_h \cdot \left( \frac{\mathbf{v} \otimes \mathbf{v}}{|\mathbf{v}|^2} \right) \cdot \nabla \varphi_h d\Omega \\ & + \sum_{e=1}^{n_{el}} \int_{\Omega_e} \chi \frac{|\mathcal{R}(\varphi_h)|}{|\nabla \varphi_h|} k_c^{\text{add}} \nabla w_h \cdot \left( \mathbf{I} - \frac{\mathbf{v} \otimes \mathbf{v}}{|\mathbf{v}|^2} \right) \cdot \nabla \varphi_h d\Omega = \int_{\Omega} w_h f d\Omega + \int_{\Gamma_N^\varphi} w_h \varphi_N d\Gamma, \end{aligned} \quad (4.55)$$

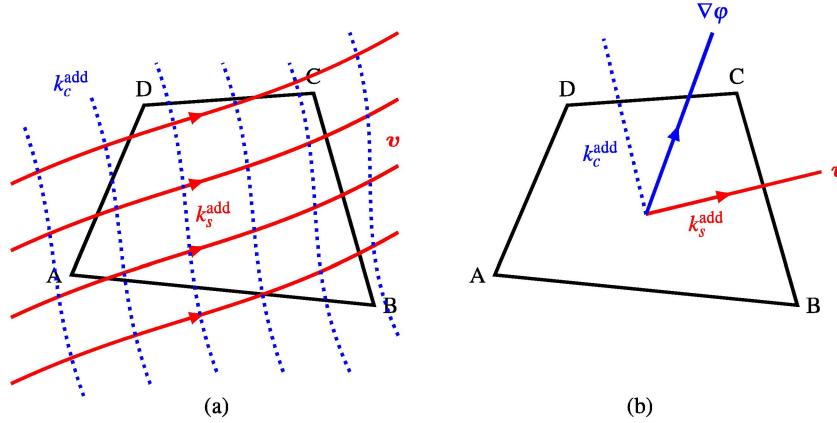
where  $k_s^{\text{add}}$  and  $k_c^{\text{add}}$  are the added diffusion in the streamline and the crosswind directions respectively. We directly extend the diffusion formula derived for the satisfaction of the positivity condition in one-dimension (Eq. (4.40)) to multi-dimensions by working with two principle directions, viz., streamline and crosswind (Fig. 4.1). Implementation of this procedure requires only minor modifications to the existing stabilized finite element codes.

*Remark 4.3.* The PPV technique adds diffusion consistently in both the streamline and crosswind directions. For the case of velocity perpendicular to the solution gradient, positivity and local boundedness is thus maintained due to the sufficient diffusion (see Fig. 4.1(b)). The effectiveness of the method has been thoroughly tested by conducting several numerical experiments in Section 4.5.

To elaborate, let us consider the streamline direction first. To maintain positivity, the following expression is adds diffusion in the streamline direction:

$$\text{Stabilization term}_s = \sum_{e=1}^{n_{el}} \int_{\Omega_e} \chi \frac{|\mathcal{R}(\varphi_h)|}{|\nabla \varphi_h|} k_s^{\text{add}} \nabla w_h \cdot \left( \frac{\mathbf{v} \otimes \mathbf{v}}{|\mathbf{v}|^2} \right) \cdot \nabla \varphi_h d\Omega, \quad (4.56)$$

where  $k_s^{\text{add}}$  is the diffusion in the streamline direction, given by



**Fig. 4.1** Sketch of the residual-based stabilization inside a generic element with ABCD sub-domain boundary: (a) Four-node quadrilateral element with convection velocity and the added diffusions  $k_s^{\text{add}}$  and  $k_c^{\text{add}}$  in the streamline and crosswind directions respectively; (b) added diffusions and solution gradient ( $\nabla \varphi$ ) shown for designing positivity. Solid (red) and dashed (blue) lines represent the streamline and crosswind directions of the convection field respectively while solid (blue) line is the solution gradient vector.

$$k_s^{\text{add}} = \max \left\{ \frac{||\mathbf{v}| - \tau|\mathbf{v}|s + \tau|\mathbf{v}||s||h}{2} - (k + \tau|\mathbf{v}|^2) + \frac{(s + \tau s|s|)h^2}{6}, 0 \right\}. \quad (4.57)$$

On the other hand, in the crosswind direction, we add the following term:

$$\text{Stabilization term}_c = \sum_{e=1}^{n_{\text{el}}} \int_{\Omega_e} \chi \frac{|\mathcal{R}(\varphi_h)|}{|\nabla \varphi_h|} k_c^{\text{add}} \nabla w_h \cdot \left( I - \frac{\mathbf{v} \otimes \mathbf{v}}{|\mathbf{v}|^2} \right) \cdot \nabla \varphi_h d\Omega, \quad (4.58)$$

where  $k_c^{\text{add}}$  is the diffusion in the crosswind direction, given by

$$k_c^{\text{add}} = \max \left\{ \frac{||\mathbf{v}| + \tau|\mathbf{v}||s||h}{2} - k + \frac{(s + \tau s|s|)h^2}{6}, 0 \right\}. \quad (4.59)$$

Here,  $|\mathbf{v}| = \sqrt{u_i u_i}$  is the magnitude of the multi-dimensional velocity vector.

*Remark 4.4.* The one-dimensional formula for  $k^{\text{add}}$  has been directly extended to multi-dimensions. Note that there is no sense of crosswind direction in one-dimension. Therefore,  $k^{\text{add}}$  in Eq. (4.40) is added along the streamline. While in multi-dimensions, the expression for streamline diffusion is same as that in one-dimension (Eq. (4.40)), and for the crosswind diffusion (Eq. (4.59)), we omit the terms related to the linear stabilization and the convection velocity. This becomes clearer if you assume the linear stabilization to be in the streamline direction (SUPG). The terms in Eq. (4.40) depicting this streamline stabilization are  $-\tau|\mathbf{v}|s|h/2$  and  $\tau|\mathbf{v}|^2$ . However, in the crosswind direction, these terms will not be present, giving Eq. (4.59). This case is elaborated as follows. Suppose that we have SUPG stabilization in one-dimension,

then the added term in the linear stabilization is

$$\sum_{e=1}^{n_{el}} \int_{\Omega^e} u \frac{dw_h}{dx} \tau \left( u \frac{d\phi_h}{dx} + s\phi_h - f \right) d\Omega. \quad (4.60)$$

The left-hand side of the stabilized expression after the finite element discretization can be written in the matrix form at the local element level as

$$\begin{bmatrix} \frac{\tau u^2}{h} - \frac{u\tau s}{2} & -\frac{\tau u^2}{h} - \frac{u\tau s}{2} \\ -\frac{\tau u^2}{h} + \frac{u\tau s}{2} & \frac{\tau u^2}{h} + \frac{u\tau s}{2} \end{bmatrix} \quad (4.61)$$

When we write the  $k^{add}$  expression for the SUPG method, these extra terms will correspond to the diffusion in the streamline direction. Therefore, the terms relating to the expressions above are omitted from the crosswind diffusion in Eq. (4.59).

*Remark 4.5.* The gradients in the solution in any direction are taken care of by the PPV technique, unlike earlier methods in the literature which added null nonlinear stabilization term when the convection velocity is perpendicular to the solution gradient.

*Remark 4.6.* The PPV technique is slightly more diffusive than that proposed in [45, 155, 156]. Neglecting the effects of the added nonlinear diffusion in the streamline direction, the method can be simplified to that of [45] by using a factor of 0.7 rather than 1 in the first term of the  $k_c^{add}$  for the convection-diffusion equation:

$$k_c^{add} = \max \left\{ \frac{0.7|\mathbf{v}|h}{2} - k, 0 \right\}. \quad (4.62)$$

Taking the constant as 1 rather than 0.7 maintains positivity throughout the computational domain. With 0.7 as the factor, a non-regular solution near the boundary for the 90° convection case is observed (Fig. 4.12(f)).

*Remark 4.7.* There are only two assumptions that have been made in the design of the method: (i) the diffusion is isotropic and (ii) higher-order derivatives in the variational form and the differential operator of the stabilization term are neglected pertaining to the use of linear and multilinear finite elements.

*Remark 4.8.* Another numerical challenge is the evaluation of the Jacobian for the implicit and nonlinear schemes due to the presence of the absolute value function. Here, we have utilized Picard fixed-point iterative procedure for the nonlinear iterations, so that  $|\mathcal{A}(\phi_h)|/|\nabla \phi_h|$  is evaluated using the known values of  $\phi_h$  at the previous nonlinear iteration.

#### 4.3.2.4 Characteristic Element Length Scale

A lack of general definition of the element length scale poses a challenge to the error analysis of the residual-based stabilization of the Galerkin formulation. It is

straightforward to select the characteristic element length in one-dimension. But, the task is very challenging in multi-dimensions, especially for highly irregular and anisotropic elements. Under generalized isoparametric transformation, the formula for the stabilization parameter  $\tau$  was shown to be [193]

$$\tau = \left[ \left( \frac{1}{\alpha \Delta t} \right)^2 + \mathbf{v} \cdot \mathbf{G} \mathbf{v} + C_I k^2 \mathbf{G} : \mathbf{G} + s^2 \right]^{-1/2}, \quad (4.63)$$

where  $C_I$  is a constant derived from inverse estimates [83] and  $\mathbf{G}$  is the element contravariant metric tensor for isoparametric mapping given by

$$\mathbf{G} = \left( \frac{\partial \boldsymbol{\xi}}{\partial \mathbf{x}} \right)^T \frac{\partial \boldsymbol{\xi}}{\partial \mathbf{x}}, \quad (4.64)$$

$\boldsymbol{\xi}$  and  $\mathbf{x}$  being the vectors of natural and global coordinates, respectively. In the case of one-dimension, the above formula for  $\tau$  reduces to Eq. (4.20). Thus, in one-dimension,

$$\mathbf{v} \cdot \mathbf{G} \mathbf{v} = \left( \frac{2u}{h} \right)^2. \quad (4.65)$$

Similarly, in multi-dimensions, we can define the element length such that one-dimensional equivalent streamline length scale is obtained. Such length scale measures have been used in the literature [49, 50, 151]:

$$h_u = \frac{2|\mathbf{v}|}{\sqrt{\mathbf{v} \cdot \mathbf{G} \mathbf{v}}}, \quad (4.66)$$

where  $h_u$  is the streamline element length and  $|\mathbf{v}| = \sqrt{u_i u_i}$ . It can be observed that the above expressions for the length scale do not consider the anisotropic character of the element, especially when there is a large directionality mismatch between the equation coefficients and the wave vector associated with subscale physics. This is the reason for the oscillations in the numerical solution for some cases when the discretization is anisotropic. A generalized Fourier analysis of the CDR equation was carried out in [173] to study this phenomenon. We highlight some key points from that analysis.

As mentioned earlier, the origin of the stabilization parameter has been attributed to the multiscale phenomenon in which the unresolved subgrid scales are modeled via residual-based stabilization. The expression for the stabilization parameter  $\tau$  is derived from the Fourier representation of the Green's function for a subscale problem at the element level, where the subscale term vanishes on the element boundary. The formula for  $\tau$  for the subscale problem can be written using the inverse Fourier transform as [173]

$$\tau = \left[ (K_i K_j k_{ij}^r + s)^2 + (K_j u_j^r)^2 \right]^{-1/2}, \quad (4.67)$$

where  $K_i$  and  $K_j$  are the components of the wave vector  $\mathbf{K}$  from the Fourier analysis,  $u'_j$  and  $k'_{ij}$  are the convection velocity and the diffusion coefficient of the transformed CDR equation in the local coordinates, i.e.,  $u'_i = u_j J_{ij}^{-T}$  and  $k'_{lm} = J_{mi}^{-T} k_{ij} J_{lj}^{-T}$  respectively, and  $J$  is the Jacobian of a finite element. By neglecting the directionality of the wave vector and writing the terms in the above expression for  $\tau$ , we get

$$K_i K_j k'_{ij} \approx ||\mathbf{K}||^2 \sqrt{k'_{ij} k'_{ij}}, \quad (4.68)$$

$$K_j u'_j \approx ||\mathbf{K}|| ||\mathbf{v}'||, \quad (4.69)$$

where  $||\cdot||$  is the invariant of the vector. The above expression is similar to Eq. (4.63) when  $s = 0$ . However, neglecting the directionality of the wave vector has consequences on an anisotropic grid with variable coefficients. Therefore, a heuristic approach was proposed in [173] to find the direction where the maximum instability occurs based on the element Peclet and Damköhler numbers. A similar definition of the element length scale is adopted here and demonstrated for anisotropic and unstructured grids in Section 4.5.3.

For generality to transient problems, all the preceding developments can be extended to the time-dependent case by considering the transient term and deriving the diffusion expressions similar to the above procedure.

#### 4.4 Convergence and Stability Analysis

In this section, a systematic analysis is carried out of the different stabilization methods for solving the one-dimensional CDR equation. Furthermore, convergence, accuracy and stability properties of the PPV method are also assessed. We assess these properties across a wide range of characteristic non-dimensional parameters. To begin, different stabilization methods are compared by analyzing the variation of the non-dimensionalized  $L^2$  error with  $Da = sh/u$  (convection-reaction case) and with  $\psi = sh^2/k$  (diffusion-reaction case). Then, the algorithmic damping and phase velocity (or algorithmic frequency) ratios are quantified for the PPV technique based on Fourier analysis to understand the stability properties. Finally, accuracy of the various methods for the steady-state convection-reaction, diffusion-reaction and CDR equations is analyzed by performing a mesh convergence study. The non-dimensionalized  $L^2$  error is computed for a fixed number of sample nodal points throughout the analysis as

$$\text{Error} = \frac{||\varphi_{\text{numerical}} - \varphi_{\text{exact}}||_2}{||\varphi_{\text{exact}}||_2}, \quad (4.70)$$

where  $\varphi_{\text{exact}}$  is the analytical value for the equation and  $||\cdot||_2$  is the  $L^2$  norm.

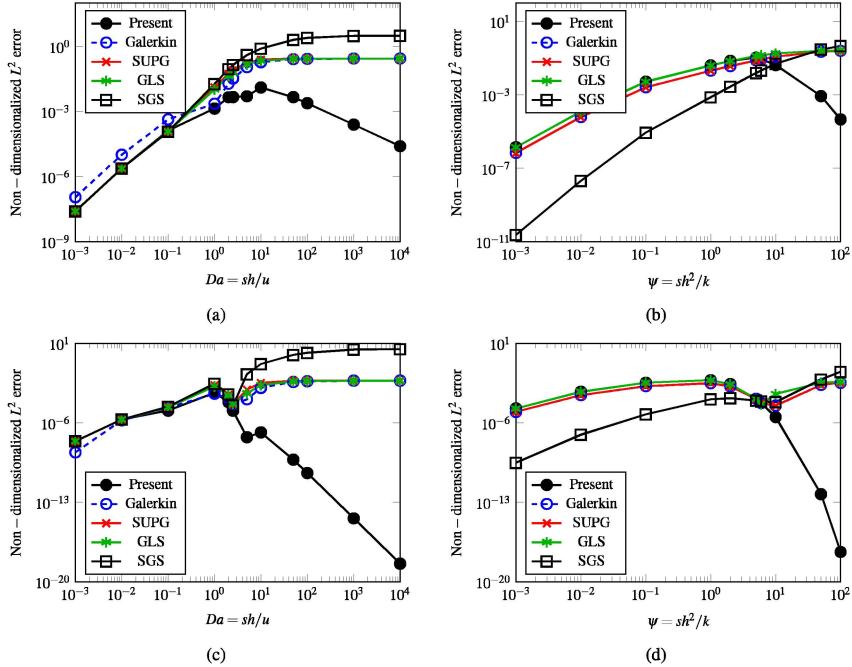
#### 4.4.1 Dependence of Error on the Non-Dimensional Parameters

For the error analysis, we consider the one-dimensional problem in a computational domain of length  $L = 1$  by varying the non-dimensional numbers while keeping the element length size to 0.1 (10 number of elements) and 0.025 (40 number of elements).

For the convection-reaction case, the left-hand side Dirichlet condition is set to 1 with  $f = 0$  and for the diffusion-reaction case, the left-hand side and right-hand side nodal values are 8 and 3 respectively with  $f = 0$ . From Figs. 4.2(a) and 4.2(c), a monotonically decreasing error is observed in the convection-dominated regime ( $Da < 1$ ) for all the techniques, while for high  $Da$ , the error from the methods apart from PPV behave like an asymptote to a constant value. On the other hand, PPV minimizes the error even in the reaction-dominated regime ( $Da > 1$ ). Moreover, the error decreases with the increase in  $Da$ . This can be attributed to the reduction in the undershoots and overshoots of the solution near the boundary layers in the case of PPV. The maximum error occurs around  $Da \approx 10$  and  $Da \approx 1$  for the coarse and fine meshes respectively.

Similarly, from Figs. 4.2(b) and 4.2(d), SGS method behaves quite well for diffusion-dominated regime ( $\psi < 1$ ) with the monotonically decreasing error as  $\psi$  decreases. An asymptotic error is observed to be reached at higher  $\psi$  for all the methods apart from PPV. The error decreases monotonically for PPV in the reaction-dominated regime ( $\psi > 1$ ). Note that the performance of the PPV is very similar to that of SUPG and GLS in the diffusion-dominated regime. The maximum error for PPV occurs around  $\psi \approx 5$  and  $\psi \approx 1$  for the coarse and fine meshes respectively. To summarize, the improvement in the error norm for PPV with increasing  $Da$  and  $\psi$  for the reaction-dominated effects is mainly due to the preservation of the positivity property and reduction of oscillations near the boundary layer.

In the middle of the regime of the non-dimensional numbers in Fig. 4.2, an increase in the computed error can be observed. This can be explained as follows. An increase in  $Da$  (or  $\psi$ ) leads to a sharper boundary layer in the exact solution of the equation. Due to the numerical discretization using finite nodes and fixed number of elements, the error depends on the ability of each node to resolve the boundary layer. At lower  $Da$  (or  $\psi$ ), the reaction effects are small, showing less numerical error. For higher  $Da$  (or  $\psi$ ), the ability of the nodes to resolve the boundary layer is tested, which introduces some error as the node may fall near the edge of the layer. For much higher non-dimensional numbers, the error decreases further due to the positivity preservation property of PPV as the solution approaches to the nodally exact one and the edge of the boundary layer recedes away from the node. The high error in the middle range of non-dimensional numbers can be reduced via an adaptive refinement as shown in the finer mesh (Fig. 4.2(c-d)) where we observe a similar trend.



**Fig. 4.2** Dependence of non-dimensionalized  $L^2$  error on non-dimensional numbers as a function of:  $Da$  for convection-reaction equation with number of elements- (a) and (c); and  $\psi$  for diffusion-reaction problem- (b) and (d). Number of elements is kept constant at 10 for (a-b) and 40 for (c-d) with a domain length  $L = 1$ .

#### 4.4.2 One- and Two-Dimensional Fourier Analysis

We continue assessing the different methods by looking into the stability and accuracy properties in this section. To accomplish this, we conduct Fourier analysis on the discretized equation without the nonlinear stabilization term and quantify the dispersion and diffusion errors.

The analysis has been carried out for the GLS and SGS methods in [112] for one- and two-dimensions. We briefly highlight the analysis here for PPV technique. Assume that the exact solution of the one-dimensional transient CDR equation can be written as a Fourier series  $\varphi = Ae^{vt+iKx}$ , where  $K$  is the spatial wave number and  $v$  denotes the evolution of the solution with time.  $v$  can be expressed as  $v = -\xi - i\omega$ ,  $\xi$  and  $\omega$  being the damping coefficient and frequency respectively and  $i = \sqrt{-1}$ . Substituting these expressions in the continuous CDR equation, one obtains

$$\xi = s + kK^2, \quad (4.71)$$

$$\omega = uK. \quad (4.72)$$

The group velocity for the continuous equation is given as  $v_g = \partial\omega/\partial K = u$ .

Similarly, the discrete solution can be written in terms of a Fourier series as  $\varphi_h(x_j, t) = Ae^{v^h t + iK^h jh}$ , where  $v^h = -\xi^h - i\omega^h$ , where  $\xi^h$  and  $\omega^h$  are the discrete counterparts of  $\xi$  and  $\omega$  termed as algorithmic damping coefficient and algorithmic frequency, respectively. It is assumed here that the temporal discretization has a negligible error for simplicity. The Fourier analysis of the PPV method in one-dimension leads to the following amplification factor:

$$v_{PPV}^h = \frac{2(\cos(Kh) - 1)(k + u^2\tau) - \frac{sh^2}{3}(\cos(Kh) + 2)(1 + |s|\tau) - ih\sin(Kh)(1 - s\tau + |s|\tau)}{\frac{h^2}{3}(\cos(Kh) + 2)(1 + |s|\tau) - ihu\tau\sin(Kh)}. \quad (4.73)$$

The above expression reduces to that of GLS method when  $s \geq 0$  and to that of SGS method when  $s < 0$ . Detailed steps of the derivation can be found in [112].

The accuracy of the PPV formulation can be analyzed by expanding  $\xi^h$  and  $\omega^h$  as follows:

$$\omega_{PPV}^h = uK + ukK^3\tau - \left(\frac{uK^5}{180}\right)h^4 + \mathcal{O}(\tau\Delta t, \Delta t^2, \tau^2, h^6), \quad (4.74)$$

$$\xi_{PPV}^h = s + kK^2 - |s|kK^2\tau + \left(\frac{kK^4}{12}\right)h^2 + \mathcal{O}(\tau\Delta t, \Delta t^2, \tau^2, h^4). \quad (4.75)$$

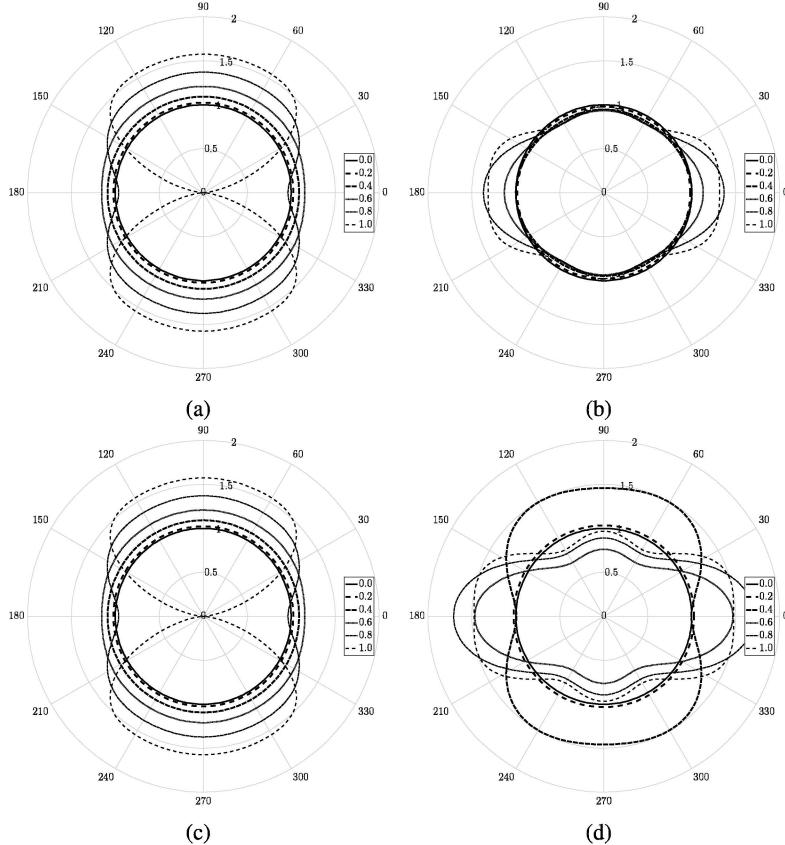
The accuracy of the algorithmic damping coefficient ( $\xi^h$ ) and the algorithmic frequency ( $\omega^h$ ) with regard to  $h$  depend on the stabilization parameter  $\tau$  when  $u \neq 0$ . Depending on the dominant phenomenon among convection, diffusion and reaction effects,  $\tau$  varies with  $h$ , according to Eq. (4.20). The algorithmic damping ratio ( $\xi^h/\xi$ ), the algorithmic frequency ratio ( $\omega^h/\omega$ ) and the group velocity ratio ( $v_g^h/c$ ) which is the ratio of the group velocity of the numerical scheme to that of the continuous differential equation, are also evaluated. It is seen that the PPV technique has relatively less damping than the SGS when  $s \geq 0$ , very similar to that of GLS. When  $s < 0$ , PPV adds appropriate damping rather than GLS, which has a larger negative damping. In this range, the PPV behaves similar to the SGS method. To summarize, the behavior of PPV will be the same as that of GLS for positive reaction and that of SGS for negative reaction. Fourier analysis of the fully discrete form of the CDR equation can be found in [112].

Extending the analysis to two dimensions, we assume the exact solution as a Fourier series as  $\varphi = Ae^{vt + i\mathbf{K} \cdot \mathbf{x}}$ , where  $\mathbf{K} = K\cos\alpha_k \mathbf{i} + K\sin\alpha_k \mathbf{j}$  and  $\mathbf{x}$  are the wave vector and the position vector respectively. It is assumed that the velocity vector  $\mathbf{v} = u\mathbf{i} + v\mathbf{j}$  has the same direction as that of the wave vector, for simplicity in the calculations. The damping coefficient and frequency for the continuous CDR equation in two-dimensions are given by

$$\xi = s + kK^2, \quad (4.76)$$

$$\omega = |\mathbf{v}|K. \quad (4.77)$$

The detailed Fourier analysis in two-dimensions is presented in [112]. In two-dimensions, the amplification factor for the PPV method can be derived as



**Fig. 4.3** Variation of phase velocity and damping ratios with  $\alpha_k$  for different  $Kh_x/\pi \in [0, 1]$  for convection-diffusion-reaction problem for the present method: Phase velocity ratio for (a) positive reaction, (c) negative reaction; Damping ratio for (b) positive reaction, (d) negative reaction.

$$\begin{aligned} v_{PPV}^h = & \frac{1}{\gamma_k h_x^2 M_x M_y (1 + |s|\tau) - i\tau h_x (u\gamma_k M_y \sin \alpha_x + v M_x \sin \alpha_y)} \times \\ & \left[ \frac{2M_x(\cos \alpha_y - 1)}{\gamma_k} (k + \tau v^2) + 2M_y(\cos \alpha_x - 1)\gamma_k (k + \tau u^2) - 2\tau u v \sin \alpha_x \sin \alpha_y \right. \\ & \left. - s\gamma_k h_x^2 M_x M_y (1 + |s|\tau) - i h_x u \gamma_k M_y \sin \alpha_x (1 - s\tau + |s|\tau) - i h_x v M_x \sin \alpha_y (1 - s\tau + |s|\tau) \right], \end{aligned} \quad (4.78)$$

where  $h_x$ ,  $\gamma_k$ ,  $\alpha_x$ ,  $\alpha_y$ ,  $M_x$  and  $M_y$  are defined in [112]. Note that the above expression can be simplified to its one-dimensional counterpart when the wave vector is considered along the coordinate directions. The phase velocity ratio is defined as  $\omega^h/\omega$  and the damping ratio as  $\xi^h/\xi$ . Figure 4.3 shows the polar plots of the phase velocity ratio and the damping ratio. Similar inferences can be made from the two-dimensional Fourier analysis as that in the one-dimensional case. When  $s \geq 0$ , SGS method is more diffusive and has more phase error, while it behaves well for  $s < 0$ .

#### 4.4.3 Mesh Convergence Study

In this section, we focus on the accuracy of the various stabilization schemes, in particular, with regard to the convergence rate. We employ the Taylor series to expand the one-dimensional finite element discretization to obtain the truncation error. Let us first consider the linear stabilized (GLS) discretized form of steady-state CDR equation:

$$\begin{aligned} & \left( -\frac{u}{2h} - \frac{k}{h^2} + \frac{s}{6} - \frac{u^2\tau}{h^2} + \frac{u\tau s}{2h} - \frac{u\tau s}{2h} + \frac{s^2\tau}{6} \right) \varphi_{h,j-1} + \left( \frac{2k}{h^2} + \frac{2s}{3} + \frac{2u^2\tau}{h^2} + \frac{2s^2\tau}{3} \right) \varphi_{h,j} \\ & + \left( \frac{u}{2h} - \frac{k}{h^2} + \frac{s}{6} - \frac{u^2\tau}{h^2} - \frac{u\tau s}{2h} + \frac{u\tau s}{2h} + \frac{s^2\tau}{6} \right) \varphi_{h,j+1} = 0. \end{aligned} \quad (4.79)$$

Expanding  $\varphi_{h,j+1}$  and  $\varphi_{h,j-1}$  around  $\varphi_{h,j}$  and rearranging, we obtain the modified equation which is solved by the numerical method [236] as

$$\begin{aligned} & u \frac{d\varphi}{dx} - k \frac{d^2\varphi}{dx^2} + s\varphi + \left( -u^2 \frac{d^2\varphi}{dx^2} + s^2\varphi \right) \tau + \frac{k}{12} \frac{d^4\varphi}{dx^4} h^2 \\ & + \left( -\frac{u^2}{12} \frac{d^4\varphi}{dx^4} + \frac{s^2}{6} \frac{d^2\varphi}{dx^2} \right) \tau h^2 + \mathcal{O}(\tau^2, h^4) = 0. \end{aligned} \quad (4.80)$$

The local truncation error can be obtained by subtracting the modified equation above with the exact differential equation, as

$$\varepsilon_{\text{GLS}} = \left( -u^2 \frac{d^2\varphi}{dx^2} + s^2\varphi \right) \tau + \frac{k}{12} \frac{d^4\varphi}{dx^4} h^2 + \left( -\frac{u^2}{12} \frac{d^4\varphi}{dx^4} + \frac{s^2}{6} \frac{d^2\varphi}{dx^2} \right) \tau h^2 + \mathcal{O}(\tau^2, h^4). \quad (4.81)$$

Now, let us observe the behavior of the truncation error under different regimes. When  $k = 0$ , i.e., in the convection-reaction regime, the coefficient of  $\tau$  becomes null using rearrangement and substitution from the convection-reaction equation. Now, the leading error is of the order  $\tau h^2$ . As  $\tau$  is of the order  $h$  when the mesh is refined, GLS is formally third-order accurate in this regime. On the other hand, note that the Galerkin method is fourth-order accurate since  $\tau = 0$  and  $k = 0$ . When  $u = 0$  in the diffusion-reaction regime, GLS is second-order accurate as  $\tau$  is of order  $h^2$  with the mesh refinement.

In a similar fashion, the truncation error for the SGS method can be derived as

$$\begin{aligned} \varepsilon_{\text{SGS}} = & \left( -u^2 \frac{d^2\varphi}{dx^2} - 2us \frac{d\varphi}{dx} - s^2\varphi \right) \tau + \frac{k}{12} \frac{d^4\varphi}{dx^4} h^2 \\ & + \left( -\frac{u^2}{12} \frac{d^4\varphi}{dx^4} - \frac{us}{3} \frac{d^3\varphi}{dx^3} - \frac{s^2}{6} \frac{d^2\varphi}{dx^2} \right) \tau h^2 + \mathcal{O}(\tau^2, h^4). \end{aligned} \quad (4.82)$$

For the this stabilization method, a third-order of accuracy is observed in the convection-reaction regime which is due to the cancellation of the coefficient of  $\tau$ . However, in the diffusion-reaction regime, a peculiar fourth-order of accuracy for

the SGS method can be seen. This can be explained if we manipulate the terms in the error such that the second-order terms get cancelled: assume that the mesh refinement could be interpreted as an increase in the diffusion effects in the equation, thus, the stabilization parameter can be approximated as

$$\tau \approx \frac{h^2}{12k}. \quad (4.83)$$

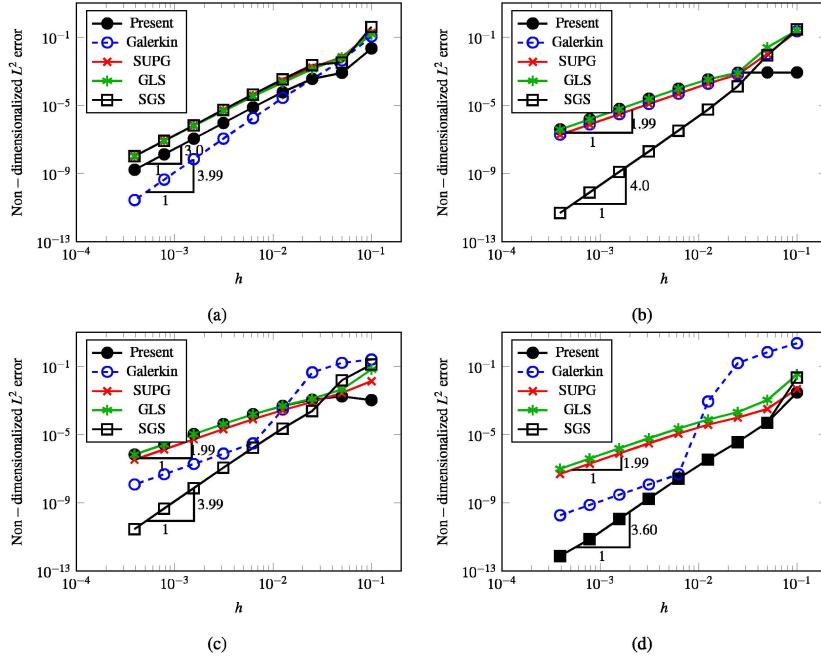
Substituting this term in  $-s^2\tau\varphi$  above and manipulating  $\frac{k}{12}\frac{d^4\varphi}{dx^4}h^2$ , we observe that these terms indeed get cancelled. Furthermore, in this case, the order of  $\tau$  is  $h^2$  leading to the order of truncation as  $\tau h^2$  making the scheme fourth-order accurate.

The truncation error for the PPV technique in the steady-state problem is given by

$$\begin{aligned} \varepsilon_{PPV} = & -\chi \frac{|\mathcal{R}(\varphi)|}{|\nabla\varphi|} k^{\text{add}} \frac{d^2\varphi}{dx^2} + \left( -u^2 \frac{d^2\varphi}{dx^2} - us \frac{d\varphi}{dx} + u|s| \frac{d\varphi}{dx} + s|s|\varphi \right) \tau \\ & + \left( \frac{k}{12} \frac{d^4\varphi}{dx^4} - \frac{1}{12} \chi \frac{|\mathcal{R}(\varphi)|}{|\nabla\varphi|} k^{\text{add}} \frac{d^4\varphi}{dx^4} \right) h^2 \\ & + \left( -\frac{u^2}{12} \frac{d^4\varphi}{dx^4} - \frac{us}{6} \frac{d^3\varphi}{dx^3} + \frac{u|s|}{6} \frac{d^3\varphi}{dx^3} + \frac{s|s|}{6} \frac{d^2\varphi}{dx^2} \right) \tau h^2 + \mathcal{O}(\tau^2, h^4). \end{aligned} \quad (4.84)$$

The PPV technique is observed to be at least second-order accurate in all the regimes. A third- and fourth-order of accuracy is noted for the convection-reaction and the propagation regime respectively. This is due to the behavior of PPV as the parent linear stabilization method, i.e., GLS when  $s \geq 0$  and SGS when  $s < 0$ ; when the residual of the equation tends to be negligible in the asymptotic convergence regime.

The observations made from the truncation analysis about the accuracy are confirmed by the mesh convergence plots in Fig. 4.4. The mesh convergence is carried out by uniformly refining the element length  $h$ , while keeping the parameters  $u$ ,  $k$  and  $s$  as constants. The mesh convergence is shown in Figs. 4.4(a), 4.4(b), 4.4(c) and 4.4(d) for the convection-reaction, the diffusion-reaction, the CDR with destruction and production cases respectively. For the CR equation,  $u = 1$ ,  $s = 50$ ,  $f = 0$  with left-hand node satisfying the Dirichlet condition of  $\varphi = 1$ ; for the DR equation,  $k = 0.01$ ,  $s = 50$ ,  $f = 0$ , left-hand and right-hand nodes satisfy  $\varphi = 8$  and  $\varphi = 3$  respectively.  $u = 4$ ,  $k = 0.01$  and  $f = 1$  with Dirichlet condition of  $\varphi = 0$  is imposed at both the extreme nodes for the CDR equation with  $s = 60$  and  $s = -6$  for destruction and production effects. The number of elements is increased by a factor of 2 from 10 to 2560 in the study. Some of the inferences from the plots are as follows. First, the measured slope of the PPV method is higher than 2 for the steady-state convection-reaction equation, as shown in Fig. 4.4(a). Second, the GLS, SUPG and PPV methods have second-order accuracy for the diffusion-reaction and the CDR equations. Third, the SGS shows an order of accuracy up to 4 for the diffusion-reaction and the CDR equations, as shown in Figs. 4.4(b)-(d).



**Fig. 4.4** Mesh convergence study for various finite element methods through dependence of non-dimensionalized  $L^2$  error as a function of uniform mesh refinement  $h$ : (a) convection-reaction; (b) diffusion-reaction; (c) CDR (destruction) and (d) CDR (production) problem. Number of elements is increased by decreasing  $h$ .

## 4.5 Numerical Results

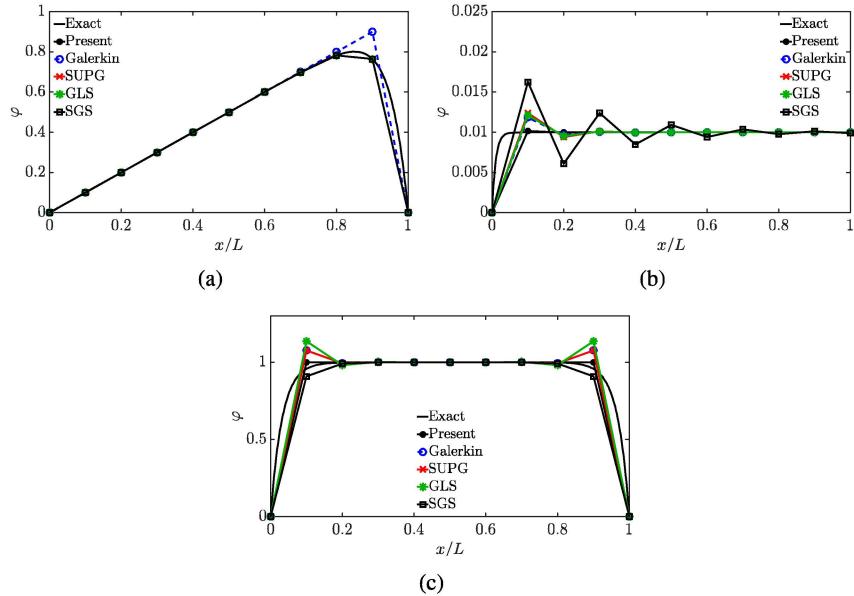
In this section, we continue assessing the different stabilization methods by conducting numerical tests for the CDR equation in one- and two-dimensions across a wide range of parameters. Moreover, we demonstrate the generality of the PPV method for unstructured and anisotropic two-dimensional meshes.

### 4.5.1 One-Dimensional Cases

We begin the numerical tests with one-dimensional steady state CDR equation. A computational domain of length  $L = 1$  is discretized into 10 linear finite elements. The parameters for the various test cases are given below:

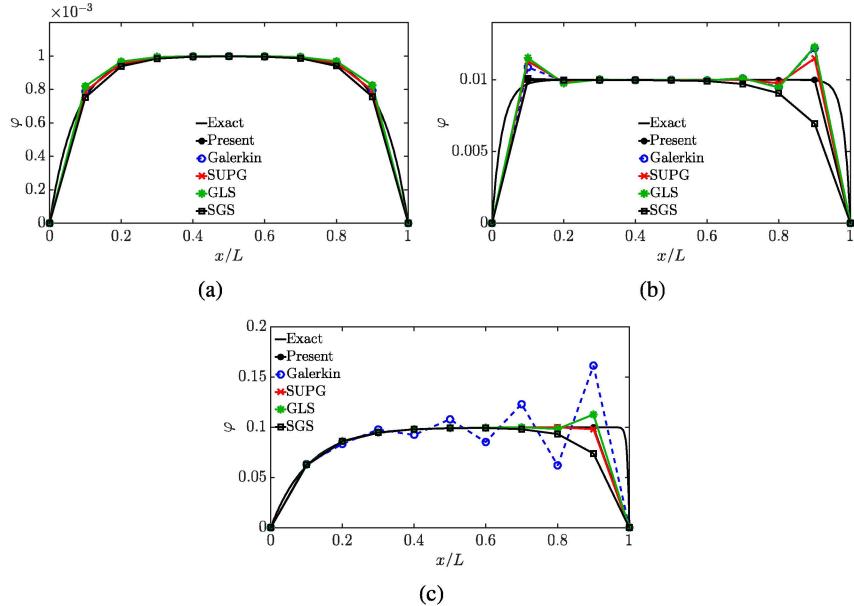
- Canonical forms
  - Convection-diffusion equation:  $Pe = 1$  with  $f = 1$  and  $(0, 0)$  (Fig. 4.5(a))

- Convection-reaction equation:  $Da = 10$  with  $f = 1$  and inlet condition = 0 (Fig. 4.5(b))
- Diffusion-reaction equation:  $\psi = 10$  with  $f = 1$  and  $(0, 0)$  (Fig. 4.5(c))



**Fig. 4.5** Variation of scalar field  $\varphi$  of the steady CDR equation in one-dimension in the three canonical regimes: (a) Convection-diffusion ( $Pe = 1$ ,  $f = 1$ ,  $(0, 0)$ ); (b) Convection-reaction ( $Da = 10$ ,  $f = 1$ , inlet Dirichlet condition = 0); (c) Diffusion-reaction ( $\psi = 10$ ,  $f = 1$ ,  $(0, 0)$ ).

- With source term ( $f \neq 0$ )
  - $Pe = 0.01$  and  $Da = 100$  with  $(0, 0)$  (Fig. 4.6(a))
  - $Pe = 1$  and  $Da = 10$  with  $(0, 0)$  (Fig. 4.6(b))
  - $Pe = 10$  and  $Da = 1$  with  $(0, 0)$  (Fig. 4.6(c))
- Without source term ( $f = 0$ )
  - $Pe = 0.1$  and  $Da = 10$  with  $(0, 1)$  (Fig. 4.7(a))
  - $Pe = 1$  and  $Da = 10$  with  $(8, 3)$  (Fig. 4.7(b))
  - $Pe = 10$  and  $Da = 10$  with  $(0, 1)$  and  $(1, 0)$  (Figs. 4.7(c) and 4.7(d))
  - $Pe = 10$  and  $Da = 0.1$  with  $(1, 0)$  (Fig. 4.7(e))
  - $Pe = 10$  and  $Da = 0.2$  with  $(8, 3)$  (Fig. 4.7(f))
  - $Pe = 10$  and  $Da = 1$  with  $(8, 3)$  (Fig. 4.7(g))
  - $Pe = 1$  and  $Da = 60$  with  $(8, 3)$  (Fig. 4.7(h))
- Production regime ( $s < 0$ )
  - Exponential regime ( $u^2 + 4ks \geq 0$ )

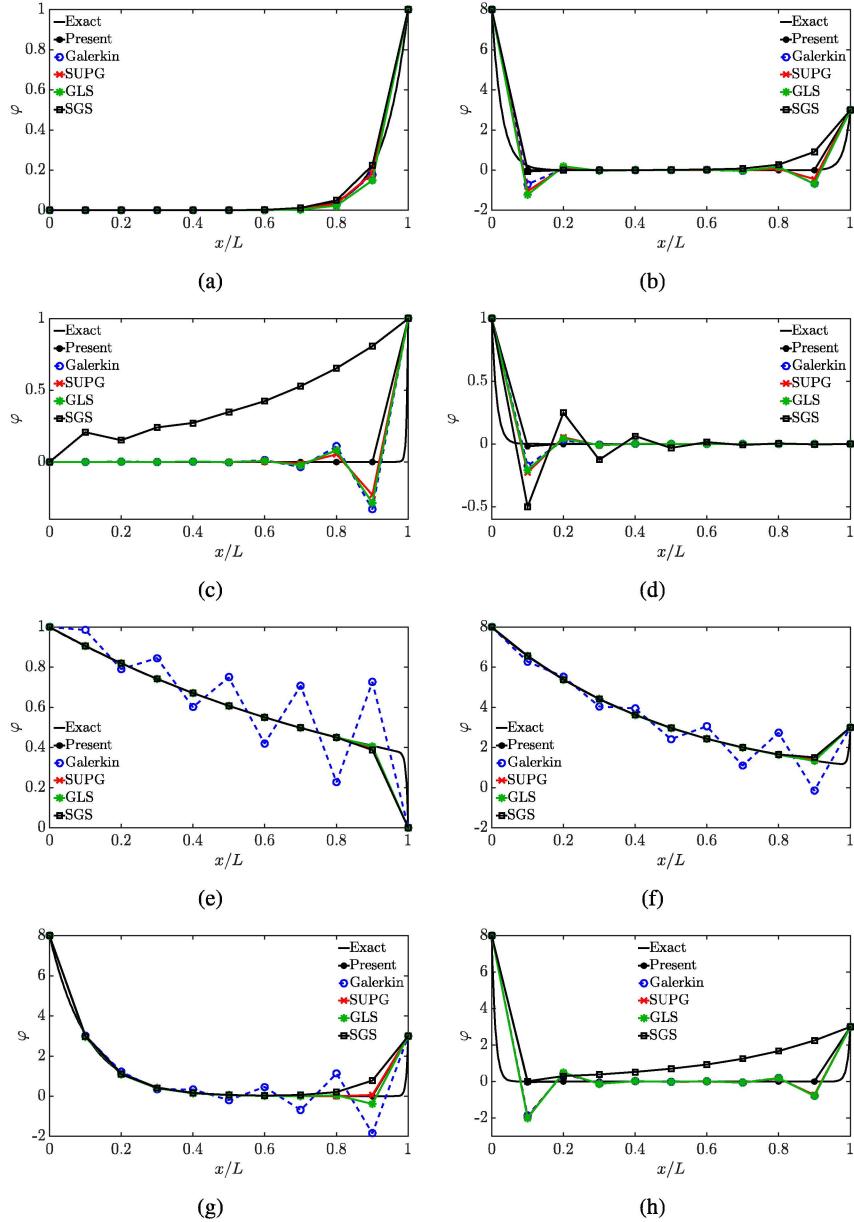


**Fig. 4.6** Solution of the scalar field  $\varphi$  of the steady CDR equation in one-dimension by setting the source term,  $f \neq 0$ : (a)  $Pe = 0.01$ ,  $Da = 100$ ,  $f = 1$ ,  $(0, 0)$ ; (b)  $Pe = 1.0$ ,  $Da = 10$ ,  $f = 1$ ,  $(0, 0)$ ; (c)  $Pe = 10$ ,  $Da = 1.0$ ,  $f = 1$ ,  $(0, 0)$ .

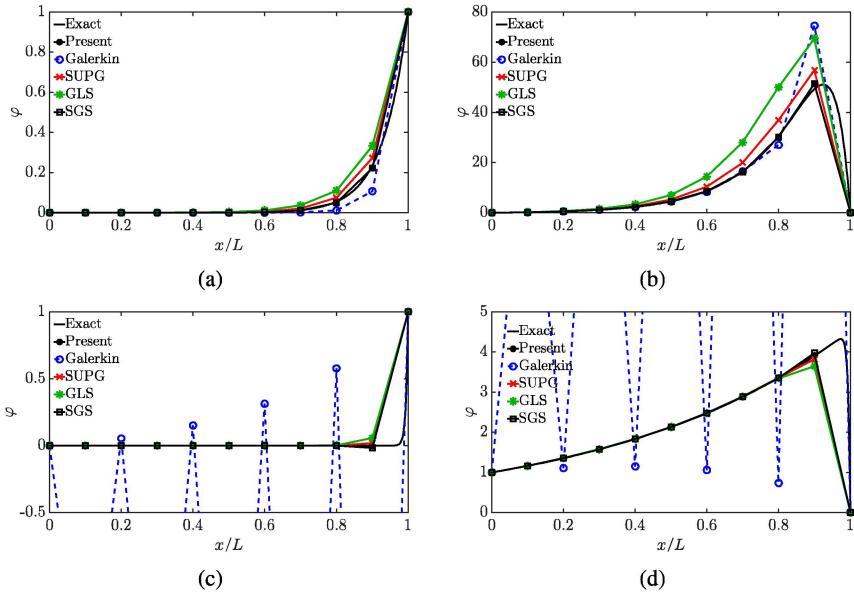
- $Pe = 1$  and  $Da = -0.4$  with  $(0, 1)$  (Fig. 4.8(a))
- $Pe = 1.5$  and  $Da = -0.5$  with  $f = 1$  and  $(0, 0)$  (Fig. 4.8(b))
- $Pe = 9$  and  $Da = -0.15$  with  $(0, 1)$  and  $(1, 0)$  (Figs. 4.8(c) and 4.8(d))
- Propagation regime ( $u^2 + 4ks < 0$ )
  - $Pe = 0.05$  and  $Da = -3$  with  $(0, 1)$  (Fig. 4.9(a))
  - $Pe = 0.1$  and  $Da = -6$  with  $(0, 1)$  (Fig. 4.9(b))
  - $Pe = 0.15$  and  $Da = -9$  with  $(0, 1)$  and  $(1, 0)$  (Figs. 4.9(c) and 4.9(d))

where  $Pe = uh/2k$ ,  $Da = sh/u$  and  $\psi = sh^2/k$  are the characteristic dimensionless quantities and the values in brackets refer to the left and right Dirichlet boundary node values respectively.

The solutions of the various cases are shown in Figs. 4.5, 4.6, 4.7, 4.8 and 4.9. The comparison of the different methods with the PPV gives a glance into its ability to preserve positivity and preclude oscillations. The other methods (SUPG, GLS and SGS) lack the enforcement of positivity and local boundedness property. The PPV technique behaves quite well in the propagation regime similar to the behavior of the SGS method (Fig. 4.9) with minimal phase error, which has been analyzed in [112]. SGS method is highly diffusive for high  $Da$  and  $Pe$  numbers (Figs. 4.7(c) and 4.7(h)). On the other hand, the Galerkin solution is very oscillatory in Figs. 4.8(c) and 4.8(d) and is partially shown.



**Fig. 4.7** Solution of the scalar field  $\phi$  of the steady CDR equation in one-dimension by setting the source term,  $f = 0$ : (a)  $Pe = 0.1, Da = 10$ ,  $(0, 1)$ ; (b)  $Pe = 1.0, Da = 10$ ,  $(8, 3)$ ; (c)  $Pe = 10, Da = 10$ ,  $(0, 1)$ ; (d)  $Pe = 10, Da = 10$ ,  $(1, 0)$ ; (e)  $Pe = 10, Da = 0.1$ ,  $(1, 0)$ ; (f)  $Pe = 10, Da = 0.2$ ,  $(8, 3)$ ; (g)  $Pe = 10, Da = 1.0$ ,  $(8, 3)$ ; (h)  $Pe = 1.0, Da = 60$ ,  $(8, 3)$ .



**Fig. 4.8** Solution of the steady CDR equation in one-dimension in the exponential regime: (a)  $Pe = 1.0$ ,  $Da = -0.4$ ,  $f = 0$ ,  $(0, 1)$ ; (b)  $Pe = 1.5$ ,  $Da = -0.5$ ,  $f = 1$ ,  $(0, 0)$ ; (c)  $Pe = 9.0$ ,  $Da = -0.15$ ,  $f = 0$ ,  $(0, 1)$ ; (d)  $Pe = 9.0$ ,  $Da = -0.15$ ,  $f = 0$ ,  $(1, 0)$ . The Galerkin solution is very oscillatory and hence is shown partially in Figs. 4.8(c) and 4.8(d).

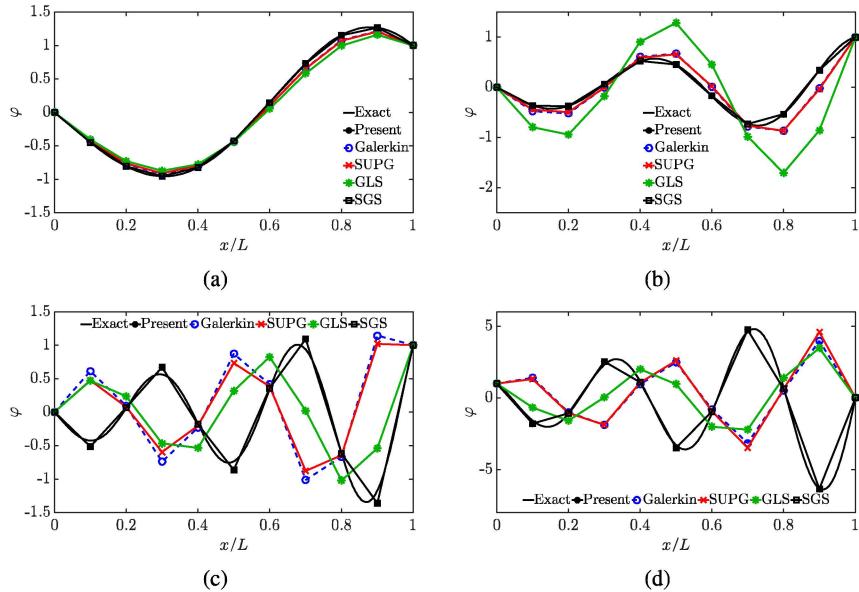
For the one-dimensional transient problem, we consider pure convection of two rectangular pulses, which was also considered in [155]. The domain is discretized using 200 linear two-node elements and the time step  $\Delta t$  is chosen as 0.002s with convection velocity  $u = 1$  and  $k = s = f = 0$ . The initial condition for the problem is

$$\varphi(x, t=0) = \begin{cases} 1, & \forall x \in [0.1, 0.2] \cup [0.3, 0.4], \\ 0, & \text{else.} \end{cases} \quad (4.85)$$

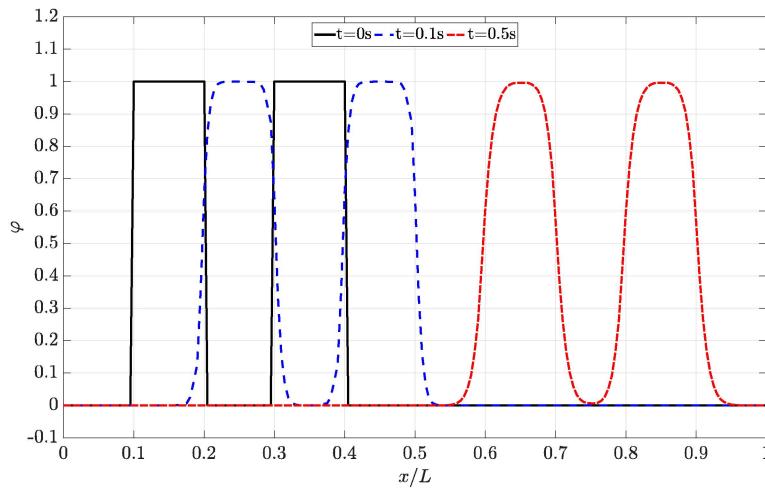
Figure 4.10 shows the evolution of the rectangular pulses at different time instances  $t = 0\text{s}$ ,  $t = 0.1\text{s}$  and  $t = 0.5\text{s}$ . No oscillations are observed during the convection phenomenon. The solution is smooth and the symmetry of the profile of the pulse is maintained at all times.

### 4.5.2 Two-Dimensional Cases

Next, we conduct numerical tests for two-dimensional steady state CDR equation. The computational domain is discretized using bilinear structured quadrilateral ele-

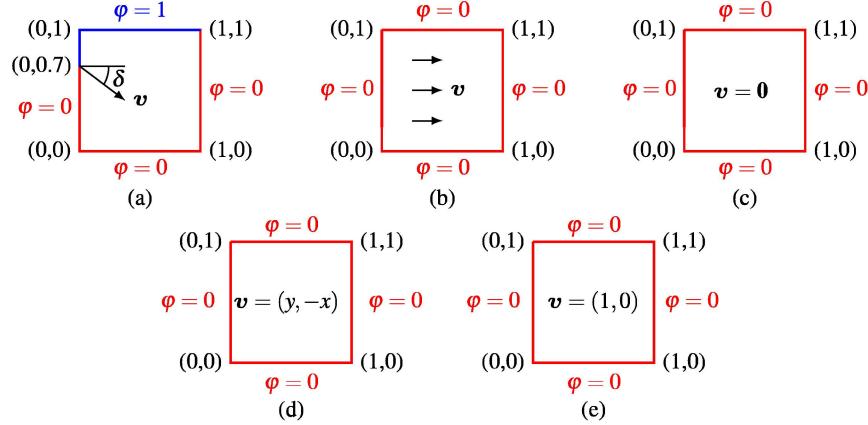


**Fig. 4.9** Solution of the steady CDR equation in one-dimension in the propagation regime: (a)  $Pe = 0.05$ ,  $Da = -3.0$ ,  $f = 0$ ,  $(0, 1)$ ; (b)  $Pe = 0.1$ ,  $Da = -6.0$ ,  $f = 0$ ,  $(0, 1)$ ; (c)  $Pe = 0.15$ ,  $Da = -9.0$ ,  $f = 0$ ,  $(0, 1)$ ; (d)  $Pe = 0.15$ ,  $Da = -9.0$ ,  $f = 0$ ,  $(1, 0)$ .



**Fig. 4.10** Solution of the transient CDR equation in one-dimension: Transient convection of two rectangular pulses,  $u = 1$ ,  $k = s = f = 0$ .

ments. Two different meshes are considered:  $20 \times 20$  and  $80 \times 80$ . The schematic of the cases with the boundary conditions are shown in Fig. 4.11 with the parameters given below:

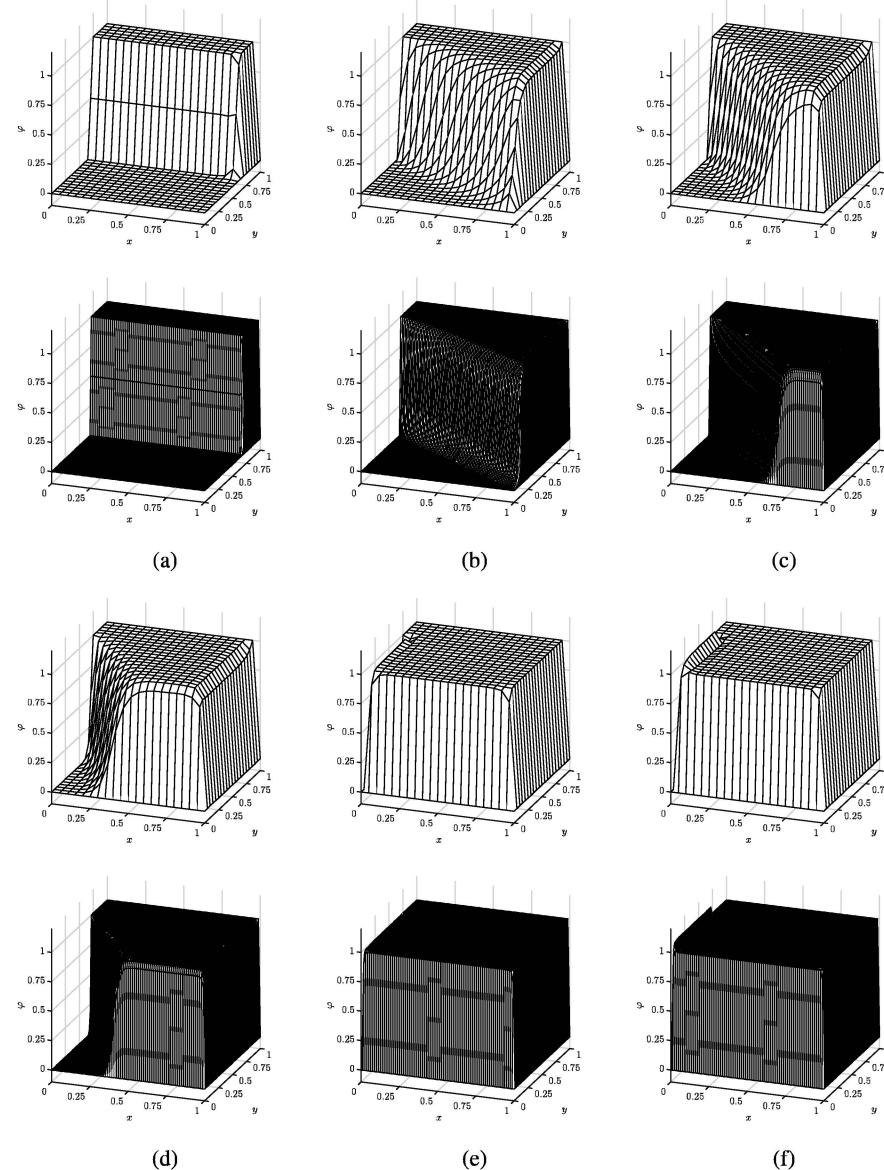


**Fig. 4.11** Schematic diagrams of the two-dimensional steady-state test cases: (a) convection-diffusion problem with unit skewed convection velocity,  $s = 0$ ,  $f = 0$ ; (b) convection-diffusion problem with constant source term,  $s = 0$ ,  $f = 1$ ; (c) diffusion-reaction problem with constant source term,  $s = 1$ ,  $f = 1$ ; (d) convection-diffusion problem with varying velocity and constant source term,  $s = 0$ ,  $f = 1$ ; (e) convection-diffusion problem with discontinuous source term,  $s = 0$ ,  $f(x \leq 0.5, y) = 1, f(x > 0.5, y) = -1$ . The diffusion coefficient ( $k$ ) is taken constant with a value of  $10^{-8}$  for all the cases.

1. Convection-diffusion problem with unit convection velocity skewed to the mesh at  $\delta = 0^\circ, 30^\circ, 45^\circ, 60^\circ$  and  $90^\circ$  with  $|\mathbf{v}| = 1$ ,  $k = 10^{-8}$ ,  $s = 0$  and  $f = 0$  (Fig. 4.12)
2. Convection-diffusion problem with constant source term.  $\mathbf{v} = (1, 0)$ ,  $k = 10^{-8}$ ,  $s = 0$  and  $f = 1$  (Fig. 4.13(a))
3. Diffusion-reaction problem with constant source term.  $\mathbf{v} = (0, 0)$ ,  $k = 10^{-8}$ ,  $s = 1$  and  $f = 1$  (Fig. 4.13(b))
4. Convection-diffusion problem with varying velocity and constant source term.  $\mathbf{v} = (y, -x)$ ,  $k = 10^{-8}$ ,  $s = 0$  and  $f = 1$  (Fig. 4.13(c))
5. Convection-diffusion with discontinuous source term.  $\mathbf{v} = (1, 0)$ ,  $k = 10^{-8}$ ,  $s = 0$  and  $f(x \leq 0.5, y) = 1, f(x > 0.5, y) = -1$  (Fig. 4.13(d))

The results for the two-dimensional steady state cases are shown in Figs. 4.12 and 4.13. The results are observed to be slightly more diffusive for the PPV method as discussed in Remark 4.6. The internal and exponential layers in the numerical solution are accurately captured with minimal oscillations, even in the case of  $90^\circ$  skewed convection. We observe some negative variation in the result for the refined case in Fig. 4.13(d) in the region where the source term  $f < 0$ . This variation is

generally manifested by all the discontinuity capturing methods based on Petrov-Galerkin formulation [46].



**Fig. 4.12** Solution of the steady-state CDR equation in two-dimensions at  $20 \times 20$  and  $80 \times 80$  mesh resolutions: Convection-diffusion problem with unit convection velocity skewed at- (a)  $0^\circ$ ; (b)  $30^\circ$ ; (c)  $45^\circ$ ; (d)  $60^\circ$ ; (e)  $90^\circ$ ; and (f)  $90^\circ$  (Codina [45]).

The transient two-dimensional test case consists of a convection of a circular bubble in a domain of  $[0, 3] \times [0, 3]$  at an angle of  $45^\circ$  with the X-axis [38, 156] with  $\mathbf{v} = (0.5, 0.5)$ ,  $k = 10^{-30}$ ,  $s = 0$  and  $f = 0$ . The domain was discretized using structured  $300 \times 300$  quadrilateral elements. The time step size was selected as  $\Delta t = 0.005$ s. The initial condition is set as

$$\varphi(\mathbf{r}, t = 0) = H(R - |\mathbf{r} - \mathbf{r}^c|), \quad (4.86)$$

where  $H()$  is the Heaviside function defined as

$$H(y) = \frac{1 + \text{sgn}(y)}{2} = \begin{cases} 0, & y < 0, \\ 0.5, & y = 0, \\ 1, & y > 0, \end{cases} \quad (4.87)$$

where  $\text{sgn}(y)$  is defined by  $\frac{y}{|y|}$ ,  $\mathbf{r}^c$  is the center of the circular bubble  $(0.5, 0.5)$ ,  $\mathbf{r}$  is the position vector in the computational domain and  $R = 0.25$ . The evolution of the bubble in time is shown in Fig. 4.14. The results show that the amplitude of the bubble is maintained in the convection process with no oscillations. The solution is visually monotone as the mesh resolves the important characteristics of the solution properly.

The errors in all the two-dimensional cases are quantified by comparing the extreme values obtained from the numerical solution with that of the exact solution in Table 4.2.

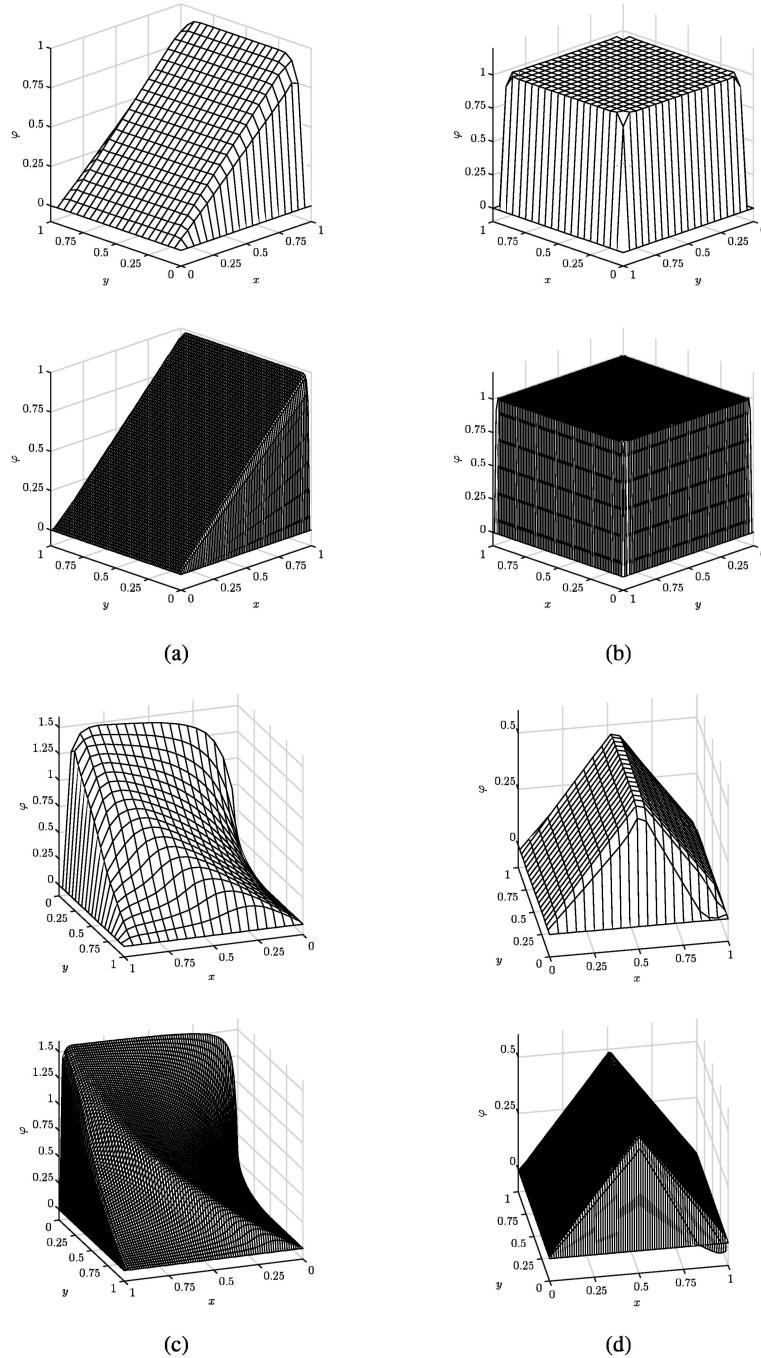
### 4.5.3 Non-Uniform Unstructured and Anisotropic Meshes

Finally, we assess the generality and robustness of the stabilization methods by conducting the numerical tests on non-uniform unstructured and anisotropic discretizations. We consider three configurations: unstructured isotropic, structured anisotropic and unstructured anisotropic quadrilateral meshes. Three cases on a computational domain of  $\Omega = [0, 1] \times [0, 1]$  are carried out:

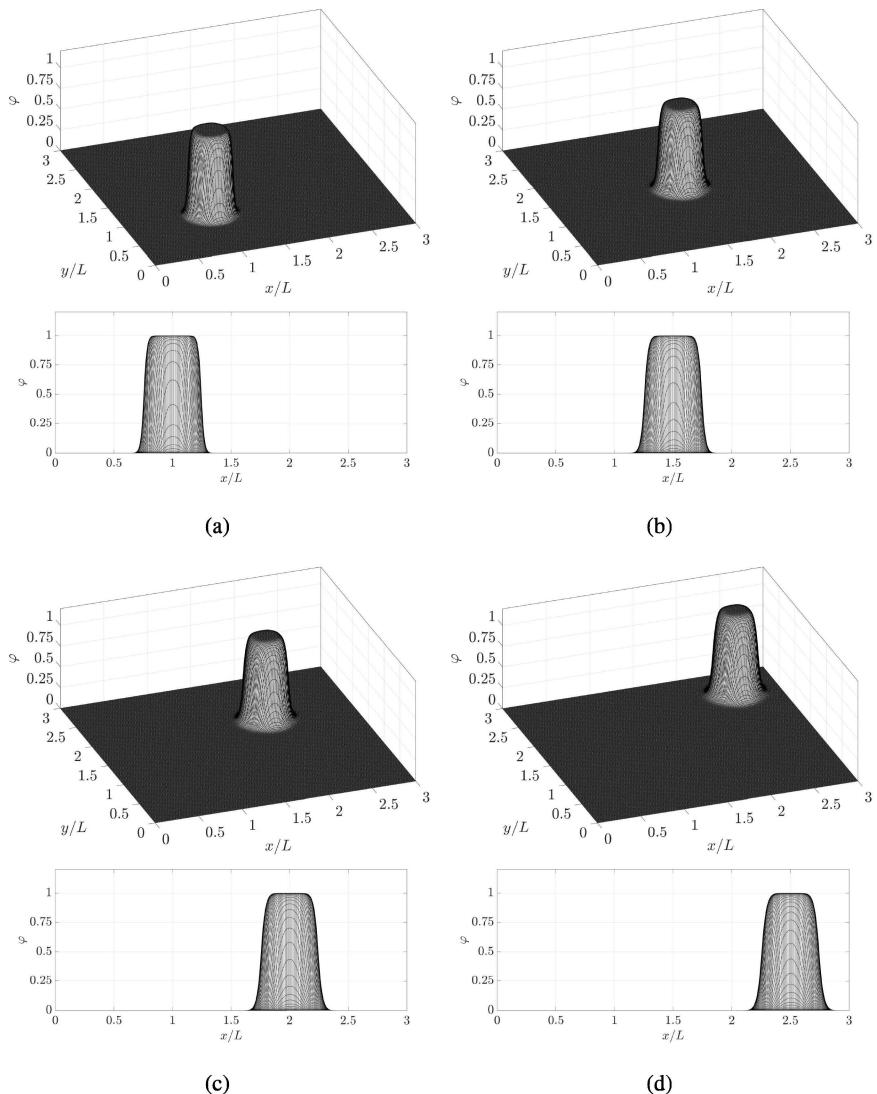
1. Skewed convection-diffusion problem at  $60^\circ$  with  $|\mathbf{v}| = 1$ ,  $k = 10^{-8}$ ,  $s = 0$  and  $f = 0$  (Fig. 4.11(a))
2. Convection-diffusion problem with constant source term with  $\mathbf{v} = (1, 0)$ ,  $k = 10^{-8}$ ,  $s = 0$  and  $f = 1$  (Fig. 4.11(b))
3. Diffusion-reaction problem with no convection with  $\mathbf{v} = (0, 0)$ ,  $k = 10^{-8}$ ,  $s = 1$ ,  $f = 1$  (Fig. 4.11(c))

As mentioned earlier, three types of mesh configurations are selected for the assessment:

- Mesh A: Unstructured quadrilateral in  $20 \times 20$  grid (Fig. 4.15(a))
- Mesh B: Structured anisotropic quadrilateral in  $10 \times 100$  grid (Fig. 4.15(b))
- Mesh C: Unstructured anisotropic quadrilateral in  $20 \times 100$  grid (Fig. 4.15(c))



**Fig. 4.13** Solution of the steady-state CDR equation in two-dimensions at  $20 \times 20$  and  $80 \times 80$  mesh resolutions: (a) Convection-diffusion problem with constant source term; (b) Diffusion-reaction problem with constant source term; (c) Convection-diffusion problem with varying velocity and constant source term; (d) Convection-diffusion problem with discontinuous source term.



**Fig. 4.14** Solution of the transient CDR equation in two-dimensions: Transient convection of a circular bubble with the convection velocity at an angle of  $45^\circ$  with the X-axis,  $k = 10^{-30}$ ,  $s = 0$ ,  $f = 0$  at- (a)  $t = 1\text{s}$ ; (b)  $t = 2\text{s}$ , (c)  $t = 3\text{s}$ ; (d)  $t = 4\text{s}$ .

**Table 4.2** Quantification of the minimum and maximum values of the variable  $\varphi$  in different test cases in two-dimensions.

Case	$\max(\varphi)$	$\max(\varphi_{\text{exact}})$	$\min(\varphi)$	$\min(\varphi_{\text{exact}})$
Fig. 4.12(a) ( $20 \times 20$ )	1.00002	1.0	$-4.06514 \times 10^{-6}$	0.0
Fig. 4.12(a) ( $80 \times 80$ )	1.00001	1.0	$-3.40949 \times 10^{-6}$	0.0
Fig. 4.12(b) ( $20 \times 20$ )	1.00165	1.0	0.0	0.0
Fig. 4.12(b) ( $80 \times 80$ )	1.00165	1.0	$-2.60600 \times 10^{-6}$	0.0
Fig. 4.12(c) ( $20 \times 20$ )	1.00077	1.0	0.0	0.0
Fig. 4.12(c) ( $80 \times 80$ )	1.00077	1.0	0.0	0.0
Fig. 4.12(d) ( $20 \times 20$ )	1.00102	1.0	0.0	0.0
Fig. 4.12(d) ( $80 \times 80$ )	1.00102	1.0	$-5.67374 \times 10^{-7}$	0.0
Fig. 4.12(e) ( $20 \times 20$ )	1.01578	1.0	0.0	0.0
Fig. 4.12(e) ( $80 \times 80$ )	1.01634	1.0	0.0	0.0
Fig. 4.13(a) ( $20 \times 20$ )	0.94999	0.95	0.0	0.0
Fig. 4.13(a) ( $80 \times 80$ )	0.98749	0.9875	0.0	0.0
Fig. 4.13(b) ( $20 \times 20$ )	1.00527	1.0	0.0	0.0
Fig. 4.13(b) ( $80 \times 80$ )	1.00532	1.0	0.0	0.0
Fig. 4.13(c) ( $20 \times 20$ )	1.50854	< 1.5708	0.0	0.0
Fig. 4.13(c) ( $80 \times 80$ )	1.55723	< 1.5708	0.0	0.0
Fig. 4.13(d) ( $20 \times 20$ )	0.50833	0.5	$-0.01208$	0.0
Fig. 4.13(d) ( $80 \times 80$ )	0.50208	0.5	$-0.04734$	0.0
Fig. 4.14(a)	1.0	1.0	0.0	0.0
Fig. 4.14(b)	1.0	1.0	0.0	0.0
Fig. 4.14(c)	0.99999	1.0	0.0	0.0
Fig. 4.14(d)	0.99999	1.0	0.0	0.0

The unstructured meshes are formed by perturbation of the coordinates of the nodes inside the domain of the structured mesh via a pseudo-random process given by

$$x_u = x_s + h_x \delta_x \text{rand}, \quad (4.88)$$

$$y_u = x_s + h_y \delta_y \text{rand}, \quad (4.89)$$

where  $(x_u, y_u)$  and  $(x_s, y_s)$  denote the nodal coordinates of the unstructured and structured meshes respectively,  $h_x$  and  $h_y$  are the mesh size for the structured mesh in  $x$ - and  $y$ -directions respectively and  $\delta_x$  and  $\delta_y$  are the mesh perturbation parameters in the respective directions with  $\text{rand}$  represents uniformly distributed random numbers in the interval  $(0, 1)$ . For mesh A, the mesh resolution  $h_x = h_y = 1/20$  and

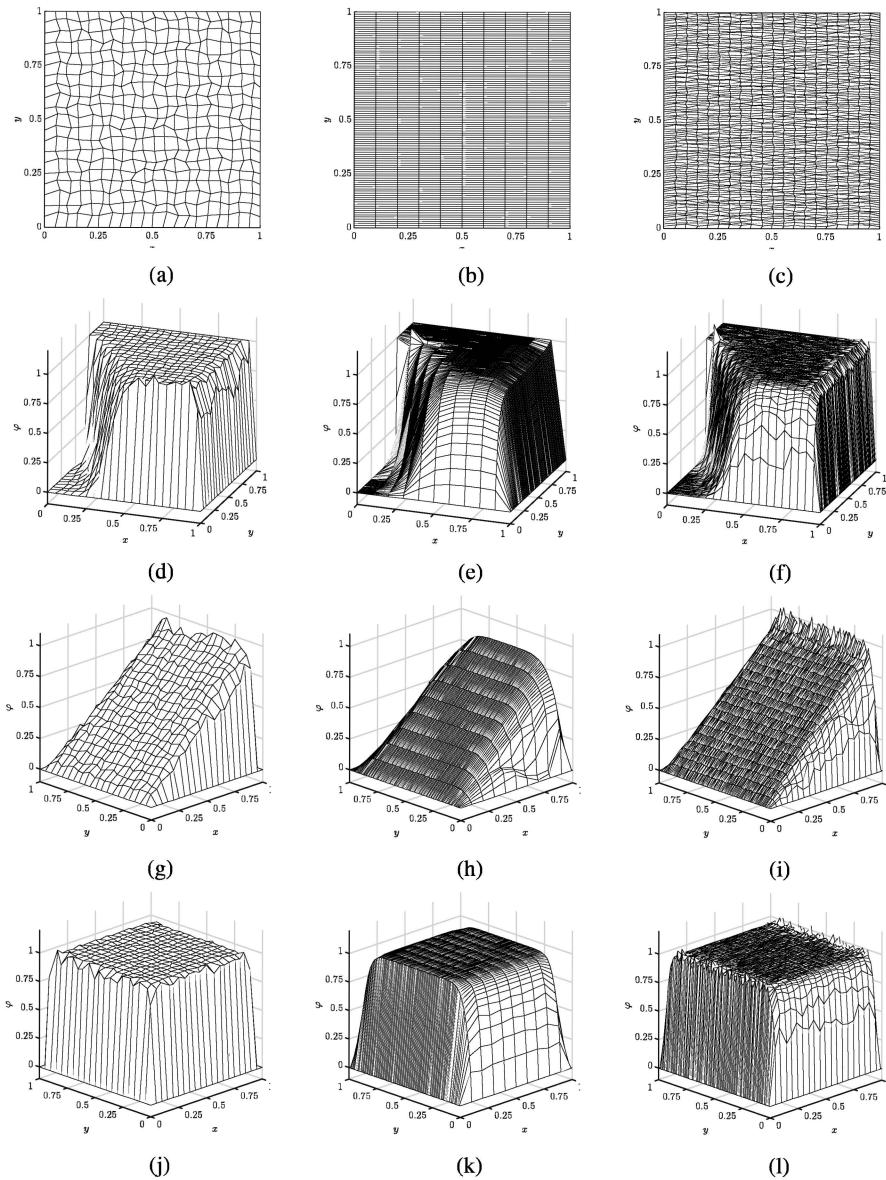
$\delta_x = \delta_y = 0.5$  is selected and for mesh C, we consider  $h_x = 1/20$ ,  $h_y = 1/100$ ,  $\delta_x = 0.2$  and  $\delta_y = 1.0$ . Figures 4.15(a-c) show the representative meshes for the various cases with the numerical solution superimposed for each case in the same figure.

The results show that the considered length scale behaves reasonably well. The spurious oscillations are reduced for the convection- and reaction-dominated cases. The internal layers are captured with reasonable accuracy in the skewed convection case with unstructured quadrilaterals (Fig. 4.15(d)) but under-resolved and highly anisotropic grids tend to have small oscillations at the extreme values of the non-dimensional numbers for the other cases. Note that using the length scale in Eq. (4.66) leads to high oscillations in the anisotropic cases. These oscillations tend to reduce by using the length scale which is derived considering the directionality of the wave vector in Eq. (4.67). The error is quantified by comparing the extreme values of the numerical solution with that of the exact solution in Table 4.3. Furthermore, the range of the non-dimensional parameters for the cases considered are summarized based on the minimum and maximum length scales obtained from Eq. (4.67) in Table 4.4.

**Table 4.3** Quantification of the minimum and maximum values of the variable  $\varphi$  in different test cases in two-dimensions: Unstructured and anisotropic meshes.

Case	$\max(\varphi)$	$\max(\varphi_{\text{exact}})$	$\min(\varphi)$	$\min(\varphi_{\text{exact}})$
Fig. 4.15(d)	1.0756	1.0	0.0	0.0
Fig. 4.15(e)	1.1111	1.0	$-1.1959 \times 10^{-4}$	0.0
Fig. 4.15(f)	1.1313	1.0	0.0	0.0
Fig. 4.15(g)	1.0012	< 1.0	0.0	0.0
Fig. 4.15(h)	0.8999	< 1.0	0.0	0.0
Fig. 4.15(i)	1.0787	< 1.0	0.0	0.0
Fig. 4.15(j)	1.0305	1.0	0.0	0.0
Fig. 4.15(k)	1.0	1.0	0.0	0.0
Fig. 4.15(l)	1.0728	1.0	0.0	0.0

This completes the discussion regarding the application of the variational stabilized finite element methods to a canonical problem of convection-diffusion-reaction equation. Error analyses are performed to quantify the accuracy, convergence and stability properties of various techniques and several numerical tests are conducted to assess the mentioned methods. Next, we continue our journey to more complex equations of fluid-structure interaction and their variational finite element discretization.



**Fig. 4.15** Three representative unstructured and anisotropic grids: (a) Mesh A, (b) Mesh B, (c) Mesh C. Results of the steady-state CDR equation in two-dimensions for skewed convection-diffusion problem at 60° (second row); Convection-diffusion problem with constant source term (third row); Diffusion-reaction problem (fourth row).

**Table 4.4** The range of non-dimensional parameters: Unstructured and anisotropic meshes.

Case	$Pe =  \mathbf{v} h/(2k)$	$Da = sh/ \mathbf{v} $	$\psi = sh^2/k$
Fig. 4.15(d)	$1.395 \times 10^6 - 3.875 \times 10^6$	0	0
Fig. 4.15(e)	$4.33 \times 10^6$	0	0
Fig. 4.15(f)	$1.755 \times 10^6 - 2.595 \times 10^6$	0	0
Fig. 4.15(g)	$1.18 \times 10^6 - 3.895 \times 10^6$	0	0
Fig. 4.15(h)	$5 \times 10^6$	0	0
Fig. 4.15(i)	$1.35 \times 10^5 - 3.01 \times 10^6$	0	0
Fig. 4.15(j)	0	$\infty$	$1.354 \times 10^5 - 6.972 \times 10^5$
Fig. 4.15(k)	0	$\infty$	$1 \times 10^6$
Fig. 4.15(l)	0	$\infty$	$1.648 \times 10^5 - 3.634 \times 10^5$

## Appendix

### A4.1 Function Spaces

We review some of the mathematical preliminaries and definitions which are helpful in formulating the weak formulation in a systematic and formal way. Consider a spatial domain  $\Omega \subset \mathbb{R}^{n_{sd}}$ , where  $n_{sd} = 1, 2$  or  $3$  based on the spatial dimensions. Let  $\Gamma$  denote the boundary to the domain. A mapping function from the domain  $\bar{\Omega} = \Omega \cup \Gamma$  to  $\mathbb{R}$ ,  $f : \bar{\Omega} \rightarrow \mathbb{R}$  is said to be of class  $C^m(\Omega)$  if all the derivatives of the function up to the order  $m$  exist and are continuous functions. While solving most of higher-order differential equations, one encounters a boundary where the first derivative becomes discontinuous leading to undefined higher derivatives. Therefore, in the variational formulation, we employ the integral form of the differential equations to reduce the burden of evaluating those higher derivatives. The topic of function spaces gives a mathematical preliminary to such space of functions which obey certain restrictions which can be helpful for circumventing the issue of undefined higher derivatives.

**Lebesgue space:** A function in the  $L^p(\Omega)$  space is considered to be Lebesgue integrable over the domain  $\Omega$  to the power  $p \in [1, \infty)$ , i.e.,

$$L^p(\Omega) = \left\{ f \left| \int_{\Omega} |f(x)|^p d\Omega < \infty \right. \right\}, \quad (4.90)$$

and is equipped with the norm

$$\|f\|_{L^p(\Omega)} = \left( \int_{\Omega} |f|^p d\Omega \right)^{1/p}. \quad (4.91)$$

One particular case of Lebesgue space is the  $L^2(\Omega)$  space which consists of functions that are square integrable over the domain. In such a case, the norm is  $\|f\|_{L^2(\Omega)} = (f, f)^{1/2}$ , where the inner product is defined as

$$(f, g) = \int_{\Omega} f g d\Omega. \quad (4.92)$$

Therefore,  $L^2(\Omega)$  is equipped with an inner product with a norm that makes it a complete metric space, a type of **Hilbert space**.

**Sobolev space:** Functions in the Sobolev space are such that they belong to  $L^p$  space and its derivatives up to a certain order  $\alpha$  also belong to  $L^p$ , i.e.,

$$W^{k,p}(\Omega) = \left\{ f \in L^p(\Omega) \left| \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \dots \partial x_{n_{sd}}^{\alpha_{n_{sd}}}} \in L^p(\Omega) \quad \forall |\alpha| \leq k \right. \right\}, \quad (4.93)$$

where  $k$  is a non-negative integer,  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_{n_{sd}}) \in \mathbb{N}^{n_{sd}}$  and  $|\alpha| = \alpha_1 + \alpha_2 + \dots + \alpha_{n_{sd}}$ .

Note that  $W^{0,2}(\Omega) = L^2(\Omega)$ . If we consider  $p = 2$ , the Sobolev space becomes a Hilbert space, i.e.,  $H^k(\Omega) = W^{k,2}(\Omega)$ . For  $k = 1$ , the Hilbert space is defined as

$$H^1(\Omega) = \left\{ f \in L^2(\Omega) \left| \frac{\partial f}{\partial x_i} \in L^2(\Omega), \quad i = 1, 2, \dots, n_{sd} \right. \right\}, \quad (4.94)$$

with the inner product and the norm respectively,

$$(f, g)_1 = \int_{\Omega} \left( fg + \sum_{i=1}^{n_{sd}} \frac{\partial f}{\partial x_i} \frac{\partial g}{\partial x_i} \right) d\Omega, \quad \|f\|_1 = \sqrt{(f, f)_1}. \quad (4.95)$$

The above function spaces can be easily extended to vector spaces. For any two vectors,  $\mathbf{u}, \mathbf{v} : \Omega \rightarrow \mathbb{R}^n$ , the Hilbert space  $H^k(\Omega)$  has the norm as,

$$\|\mathbf{u}\|_k = \left( \sum_{i=1}^n \|u_i\|_k^2 \right)^{1/2}. \quad (4.96)$$

#### A4.2 Stabilization Methods in the Literature

The concept of the stabilization methods is to provide some kind of upwinding phenomenon in the numerical scheme to add stability and accuracy. The upwinding function, introduced in [93] was noticed to cause an excessive crosswind diffusion and was limited only in the streamline direction. A streamline upwind Petrov-Galerkin (SUPG) method was suggested by [34] which introduces upwinding in the convection, transient as well as source terms. The SUPG method does not prevent overshooting and undershooting of the solution near the sharp boundary and internal layers in multi-dimensions. The numerical solution may be oscillatory or unbounded

owing to the loss of conservation at the discrete level and the existing discontinuities. Satisfying the discrete maximum principle (DMP) preserves this positivity behavior by averting any maxima or minima inside the computational domain. The finite element local matrix was constructed to be non-negative and satisfy this DMP condition in [152], but the scheme was restricted to linear triangular elements. In [176], a monotone method with conservation properties was proposed where interpolation of the element geometry and flow rates led to the approximation of the convection term. This method was also limited to bilinear quadrilateral elements.

Apart from the linear methods mentioned above, nonlinear methods satisfying the DMP condition and circumventing the Godunov's theorem [135] (which limits the accuracy of a linear monotone method to first-order) for finite difference approximation were put forward. In the finite element context, a nonlinear discontinuity capturing term was introduced in the direction of the solution gradient [99] which satisfied the DMP. Some other modifications to the variational finite element formulation were proposed in [70, 58]. To improve convergence and stability of the method with accuracy, methods based on least-squares framework such as Galerkin/least-squares (GLS) and gradient Galerkin/least-squares (GGLS) method were introduced in [97] and [68] respectively. Nonlinearity to the GLS method was introduced by a discontinuity capturing term in [193], with a simplified formula for the stabilization parameter. Further development in [45] suggested the use of cross-wind diffusion instead of isotropic diffusion for the nonlinear term, providing relatively accurate and improved convergence in the results. A review and comparison of the various linear stabilization methods such as SUPG, space-time Galerkin/least-squares (ST-GLS), subgrid-scale (SGS), characteristic Galerkin (CG) and Taylor-Galerkin (TG) was carried out in [47].

The origin of the stabilization parameter was attributed to the subgrid-scale phenomena in [94, 96]. The concept lies in the fact that across the spectrum of scales to be captured by the discretized method, the large scales are resolved by the method and the small unresolvable scales are modeled by Fourier representation of the Green's function of the subscale problem. The residual-based variational multiscale (RBVMS) method arises from the concept mentioned above with its application to the Navier-Stokes equations in [22, 4]. The stabilization parameter was further derived using the Taylor series and the concept of flow balance in a finite domain using the finite increment calculus (FIC) method in [162, 163, 165, 164], which gave another interpretation to its origins. Furthermore, the concept of flux corrected schemes (FEM-FCT) and total variational diminishing schemes (FEM-TVD) in the finite element context proposed in [27, 244, 131] added an anti-diffusive flux to the first-order accurate method (discrete upwinding), to capture the discontinuities and shocks. These type of methods are referred as high resolution methods for finite-difference/finite-volume approximations [85].

The stabilization methodology was extended to the transport equation consisting of the reaction effects leading to SUPG with diffusion for reaction-dominated regions (SUPG+DRD), SUPG with centered Petrov-Galerkin (SU+C)PG and Galerkin/least-squares/gradient least-squares (GLS/GLS) methods in [210], [102] and [84] respectively. The transient convection-diffusion equation was shown to behave as a trans-

port equation with reaction effects in [101]. For the CDR equation, the stabilization parameter  $\tau$  was derived while satisfying the DMP in [47], however the method did not generate a non-negative matrix in multi-dimensions. Methods based on the subgrid-scale phenomena were proposed for the reaction-dominated regimes as algebraic subgrid-scale method (SGS), unusual stabilized finite element method (US-FEM) and its variants and methods employing the Green's function approach in [48], [69] and [87, 86] respectively. Stabilization for the CDR equation on distorted meshes employing the VMS technique was carried out in [173]. Nonlinear finite element methods using the discrete upwind operator to satisfy the DMP condition for convection- and reaction-dominated regimes were proposed in [155, 156, 112].