CONNECTIONS BETWEEN FINITE DIFFERENCE AND FINITE ELEMENT APPROXIMATIONS FOR A CONVECTION-DIFFUSION PROBLEM

CRISTINA BACUTA and CONSTANTIN BACUTA

Communicated by Gabriela Marinoschi

We consider a model convection-diffusion problem and present useful connections between the finite differences and finite element discretization methods. We introduce a general upwinding Petrov–Galerkin discretization based on bubble modification of the test space and connect the method with the general upwinding approach used in finite difference discretization. We write the finite difference and the finite element systems such that the two corresponding linear systems have the same stiffness matrices, and compare the right-hand side load vectors for the two methods. This new approach allows for improving well-known upwinding finite difference methods and for obtaining new error estimates. We prove that the exponential bubble Petrov–Galerkin discretization can recover the interpolant of the exact solution. As a consequence, we estimate the closeness of the related finite difference solutions to the interpolant. The ideas we present in this work, can lead to building efficient new discretization methods for multidimensional convection dominated problems.

AMS 2020 Subject Classification: 35K57, 65N06, 65N12, 65N22, 65N30, 74S05.

Key words: finite difference, finite element, Petrov–Galerkin, upwinding, convection dominated problem, singularly perturbed problems.

1. INTRODUCTION

We start with the model of a singularly perturbed convection diffusion problem: Find $u = u(x)$ on $[0,1]$ such that

$$\begin{cases} -\varepsilon u''(x) + \kappa\, u'(x) = f(x), \quad 0 < x < 1 \\ u(0) = 0,\ u(1) = 0, \end{cases} \tag{1}$$

where ε and κ are positive constants. In this paper, we consider the convection dominated case, i.e. $\varepsilon \ll 1$. Here, the function f is given and assumed to be continuous on $[0,1]$. Without loss of generality, we will further assume that $\kappa = 1$.

The authors' work on this material was supported by NSF-DMS 2011615.

The model problem (1) and its multi-dimensional variants arise when solving heat transfer problems in thin domains, as well as when using small step sizes in implicit time discretizations of parabolic convection diffusion type problems, see [21]. The solutions to these problems are characterized by boundary layers, see e.g., [15, 18, 22, 27, 29]. Approximating such solutions poses numerical challenges due to the ε-dependence of the stability constants and of the error estimates. There is a tremendous amount of literature addressing these types of problems, see e.g. [18, 22, 26, 27, 29]. The goal of this paper is to use connections between upwinding Finite Differences (FD) and certain Finite Element (FE) discretizations of the model convection diffusion problem (1), to improve the performance of the upwimding FD methods, and to find new error estimates for both methods. We introduce a general upwinding FE Petrov–Galerkin (PG) discretization based on bubble modification of the test space. The test space is modified by using translations of a generating bubble function.

For studying stability error estimates and connection with FD methods, we use the concept of optimal trial norm, as presented in [1, 2, 12, 13, 15, 17, 24]. We write the finite difference and the finite element systems for uniformly distributed nodes such that the two corresponding linear systems have the same stiffness matrices, and compare the Right-Hand Side (RHS) load vectors for the two methods. The same technique is applied for the corresponding variational formulations of the FE and FD methods by finding a common bilinear form and by comparing the RHS functionals. We emphasize that any upwinding FD method can be deduced from an FE PG method by carefully selecting the generating bubble function and a quadrature to approximate the RHS dual vector of the FE PG system. The approach allows for improving the performance of known upwinding FD approaches. In addition, we investigate a particular PG method based on an exponential generating bubble function and prove that the method recovers the interpolant of the exact solution. This leads to new error estimates for the corresponding upwinding FD method.

The rest of the paper is organized as follows. We review the upwinding FD method in Section 2, and the FE discretization together with the concept of optimal trial space in Section 3. We introduce a general upwinding Petrov–Galerkin discretization method and relate it with the upwinding FD method in Section 4. In Section 5 and Section 6, we define particular test spaces based on quadratic bubbles and exponential type bubbles, respectively, and connect the new PG methods with known upwinding FD methods.

2. STANDARD FINITE DIFFERENCE DISCRETIZATION

In this section, we review the standard upwinding FD discretization of (1) on $[0,1]$ on uniform meshes. We divide the interval into n subintervals using the uniformly distributed nodes $x_j = hj$, with $j = 0, 1, \ldots, n$ and $h = \frac{1}{n}$, and consider the second order finite difference approximation for $u''(x_j)$ and $u'(x_j)$ at the nodes x_{j-1}, x_j, and x_{j+1}, to obtain the linear system:

$$(2) \qquad \begin{cases} u_0 = 0 \\ -\varepsilon \dfrac{u_{j-1} - 2u_j + u_{j+1}}{h^2} + \dfrac{-u_{j-1} + u_{j+1}}{2h} = f_j \quad j = \overline{1, n-1}, \\ u_n = 0, \end{cases}$$

where $f_j = f(x_j)$. Multiplying the generic equation in (2) by h, gives

$$(3) \qquad \begin{cases} u_0 = 0 \\ \varepsilon \dfrac{-u_{j-1} + 2u_j - u_{j+1}}{h} + \dfrac{-u_{j-1} + u_{j+1}}{2} = h\, f_j \quad j = \overline{1, n-1}. \\ u_n = 0. \end{cases}$$

Since the convection coefficient $\kappa = 1 > 0$, the standard FD *upwinding method* for discretizing (1), requires the *backward difference method* for approximating $u'(x_j)$ and leads to the system

$$(4) \qquad \begin{cases} u_0 = 0 \\ \varepsilon \dfrac{-u_{j-1} + 2u_j - u_{j+1}}{h} + (-u_{j-1} + u_j) = h\, f_j \quad j = \overline{1, n-1}. \\ u_n = 0. \end{cases}$$

Using that

$$u_j - u_{j-1} = \frac{-u_{j-1} + u_{j+1}}{2} + \frac{-u_{j-1} + 2u_j - u_{j+1}}{2},$$

the system (4) becomes

$$(5) \qquad \begin{cases} u_0 = 0 \\ \varepsilon \left(1 + \dfrac{h}{2\varepsilon}\right) \dfrac{-u_{j-1} + 2u_j - u_{j+1}}{h} + \dfrac{-u_{j-1} + u_{j+1}}{2} = h\, f_j \\ u_n = 0. \end{cases}$$

The quantity $\frac{h}{2\varepsilon}$ is known as *the local Peclet number* and is denoted by

$$\mathbb{P}e = \frac{h}{2\varepsilon}.$$

According to Section 12.5 of [26], the upwinding scheme (4) can be viewed as a centered difference scheme for the convection term with a correction for the coefficient of the diffusion term by $\varepsilon \cdot \mathbb{P}e$. The correction process is known as adding *artificial diffusion* or *numerical viscosity*. As presented in [26, 28], more general discretization based on *artificial diffusion* can be written as follows:

$$\left\{ \begin{array}{rl} u_0 & = 0 \\ \varepsilon_h \, \dfrac{-u_{j-1} + 2u_j - u_{j+1}}{h} + \dfrac{-u_{j-1} + u_{j+1}}{2} & = h\, f_j \\ u_n & = 0, \end{array} \right. \tag{6}$$

where, for a smooth function $\Phi : (0, \infty) \to (0, \infty)$ with $\lim\limits_{t \to 0} \Phi(t) = 0$,

$$\varepsilon_h = \varepsilon \left(1 + \Phi(\mathbb{P}e)\right).$$

We note that $\Phi(\mathbb{P}e) = \mathbb{P}e$ corresponds to the standard upwinding method (5).

For the general upwinding case, from (6), we obtain the system

$$\left(\frac{\varepsilon_h}{h} S + C\right) U = F_{fd}, \tag{7}$$

where $U, F \in \mathbb{R}^{n-1}$ and $S, C \in \mathbb{R}^{(n-1)\times(n-1)}$ with:

$$U := \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_{n-1} \end{bmatrix}, \quad F_{fd} := h \begin{bmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_{n-1}) \end{bmatrix} \text{ and}$$

$$S := \begin{bmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 2 \end{bmatrix}, \quad C := \frac{1}{2} \begin{bmatrix} 0 & 1 & & & \\ -1 & 0 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 0 & 1 \\ & & & -1 & 0 \end{bmatrix}. \tag{8}$$

Using the "tridiagonal notation", we have that

$$S = \text{tridiag}(-1, 2, -1), \; C = \text{tridiag}\Big(-\frac{1}{2}, 0, \frac{1}{2}\Big),$$

and the matrix of the finite difference system (7) is

$$M_{fd} = \text{tridiag}\Big(-\frac{\varepsilon_h}{h} - \frac{1}{2}, \; \frac{2\,\varepsilon_h}{h}, \; -\frac{\varepsilon_h}{h} + \frac{1}{2}\Big). \tag{9}$$

In Section 4, we will see that the general FD discretization (6), leading to (7), relates to a Petrov–Galerkin method with a bubble test space.

3. FINITE ELEMENT LINEAR VARIATIONAL FORMULATION AND DISCRETE OPTIMAL TRIAL NORM

For the finite element discretization, we will use the following notation:

$$a_0(u,v) = \int_0^1 u'(x)v'(x)\,dx,\ (f,v) = \int_0^1 f(x)v(x)\,dx, \text{ and}$$
$$b(v,u) = \varepsilon\, a_0(u,v) + (u',v) \text{ for all } u, v \in V = Q = H_0^1(0,1).$$

A variational formulation of (1), with $\kappa = 1$, is: Find $u \in Q := H_0^1(0,1)$ such that

$$b(v,u) = (f,v), \text{ for all } v \in V = H_0^1(0,1). \tag{10}$$

The existence and uniqueness of the solution of (10) is well known, see e.g., [8]–[11, 16, 19, 25].

3.1. Standard discretization with $C^0 - P^1$ test and trial spaces

We divide the interval $[0,1]$ into n equal length subintervals using the nodes $0 = x_0 < x_1 < \cdots < x_n = 1$ and denote $h := x_j - x_{j-1} = 1/n$. For the above uniformly distributed nodes, we define the corresponding finite element discrete space $\mathcal{M}_h$ as the subspace of $Q = H_0^1(0,1)$, given by

$$\mathcal{M}_h = \big\{v_h \in Q \mid v_h \text{ is linear on each } [x_j, x_{j+1}]\big\},$$

i.e., $\mathcal{M}_h$ is the space of all *continuous piecewise linear functions* with respect to the given nodes, that *are zero at* $x = 0$ *and* $x = 1$. We consider the nodal basis $\{\varphi_j\}_{j=1}^{n-1}$ with the standard defining property $\varphi_i(x_j) = \delta_{ij}$. We couple the above discrete trial space with the discrete test space $V_h := \mathcal{M}_h$. Thus, the discrete variational formulation of (10) is: Find $u_h \in \mathcal{M}_h$ such that

$$b(v_h, u_h) = (f, v_h), \text{ for all } v_h \in V_h. \tag{11}$$

We look for $u_h \in V_h$ with the nodal basis expansion

$$u_h := \sum_{i=1}^{n-1} u_i \varphi_i, \text{ where } u_i = u_h(x_i).$$

If we consider the test functions $v_h = \varphi_j, j = 1, 2, \cdots, n-1$ in (11), we obtain the following linear system

$$\Big(\frac{\varepsilon}{h} S + C\Big)\, U = F_{fe}, \tag{12}$$

where $S, C \in \mathbb{R}^{(n-1)\times(n-1)}$ are given in (8), and

$$U := \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_{n-1} \end{bmatrix}, \quad F_{fe} := \begin{bmatrix} (f,\varphi_1) \\ (f,\varphi_2) \\ \vdots \\ (f,\varphi_{n-1}) \end{bmatrix}.$$

3.2. **Optimal discrete trial norms**

For $V = Q = H_0^1(0,1)$, we consider the standard inner product given by $a_0(u,v) = (u,v)_V = (u',v')$, and let $\mathcal{M}_h, V_h$ be the standard space of continuous piecewise linear functions

$$\mathcal{M}_h = V_h = \text{span}\{\varphi_1, \cdots, \varphi_{n-1}\}.$$

For the purpose of error analysis, on V and V_h, we consider the standard norm induced by $a_0(\cdot,\cdot)$, i.e, $|v|^2 := a_0(v,v)$, but on $\mathcal{M}_h$, we will introduce a different norm. On $V_h \times \mathcal{M}_h$, we define the bilinear form

$$b_d(v_h, u_h) = d\, a_0(u_h, v_h) + (u_h', v_h) \text{ for all } u_h \in \mathcal{M}_h, v_h \in V_h, \tag{13}$$

where $d = d_{\varepsilon,h}$ is a constant, that might depend on h and ε and is motivated by the *upwinding Petrov–Galerkin method* introduced in Section 4.1.

The discrete optimal trial norm on $\mathcal{M}_h$ is defined by

$$\|u_h\|_{*,h} := \sup_{v_h \in V_h} \frac{b_d(v_h, u_h)}{|v_h|}. \tag{14}$$

An explicit representation of this norm is established in [3]. We have

$$\|u_h\|_{*,h}^2 = d^2 |u_h|^2 + |u_h|_{*,h}^2, \tag{15}$$

where,

$$|u|_{*,h}^2 = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{1}{h} \int_{x_{i-1}}^{x_i} u(x)\, dx \right)^2 - \left(\int_0^1 u(x)\, dx \right)^2. \tag{16}$$

We note that $|\cdot|_{*,h}$ is a semi-norm on V_h, since we can have $|u_h|_{*,h} = 0$ for any non-zero function $u_h \in \mathcal{M}_h$ such that

$$\int_{x_{i-1}}^{x_i} u_h(x)\, dx = \frac{1}{n} \int_0^1 u_h(x)\, dx, \ i = 1, 2, \cdots, n.$$

In particular, for $n = 2m$, such a function is $u_h = \varphi_1 + \varphi_3 + \cdots + \varphi_{2m-1}$. Using the Cauchy–Schwarz inequality, we can also check that $|u|_{*,h} \le \|u\|$. Indeed,

$$|u|_{*,h}^2 \le \frac{1}{n} \sum_{i=1}^{n} \left(\frac{1}{h} \int_{x_{i-1}}^{x_i} u(x)\, dx \right)^2 \le \sum_{i=1}^{n} \int_{x_{i-1}}^{x_i} u^2(x)\, dx = \|u\|^2. \tag{17}$$

The above estimates together with the formula (15) suggest that the discrete optimal trial norm $\|\cdot\|_{*,h}$ could be a weak norm on $\mathcal{M}_h$ if d is very small.

The optimal trial norm helps with stability estimates in the following sense. If we consider the problem: Find $u_h \in \mathcal{M}_h$ such that

$$b_d(v_h, u_h) = F_h(v_h), \text{ for all } v_h \in V_h = \mathcal{M}_h, \tag{18}$$

where $F_h : V_h \to \mathbb{R}$ is a linear functional on V_h, then according to definition (14), we have

$$\|u_h\|_{*,h} := \sup_{v_h \in V_h} \frac{b_d(v_h, u_h)}{|v_h|} = \sup_{v_h \in V_h} \frac{F_h(v_h)}{|v_h|} := \|F_h\|_{V_h^*}. \tag{19}$$

Remark 3.1. In particular, if F_1 and F_2 are two linear functionals on V_h and we solve for $u_h^1, u_h^2 \in \mathcal{M}_h$ such that

$$\begin{aligned} b_d(v_h, u_h^1) &= F_1(v_h), \text{ for all } v_h \in V_h = \mathcal{M}_h, \text{and} \\ b_d(v_h, u_h^2) &= F_2(v_h), \text{ for all } v_h \in V_h = \mathcal{M}_h, \end{aligned} \tag{20}$$

then

$$\|u_h^2 - u_h^1\|_{*,h} = \|F_2 - F_1\|_{V_h^*}. \tag{21}$$

As an application, we can have u_h^1 be the FD approximation and u_h^2 be the FE approximation of problem (1). In this case, we can estimate the difference between the two solutions in the $\|\cdot\|_{*,h}$ norm by finding an upper bound for $\|F_2 - F_1\|_{V_h^*}$.

4. THE PETROV–GALERKIN METHOD WITH BUBBLE TYPE TEST SPACE

For improving the stability and approximability of the standard linear finite element approximation for solving (10), various Petrov–Galerkin discretizations were considered, see e.g., [3, 6, 4, 5, 14, 23, 28, 29]. In this section, we introduce a general class of upwinding PG discretizations based on a bubble modification of the standard $C^0 - P^1$ test space. The idea is to define V_h by adding to each φ_j, a pair of polynomial bubble functions. According to Section 2.2.2 in [29], this idea was first suggested in [20] and used in the same year in [14] with quadratic bubble modification. The method is known in literature as *upwinding PG method*, according to Section 2.2.2 in [29], or *upwinding finite element method*, according to Section 2.2 in [28].

Besides a more general approach of the method, we discover an equivalent variational reformulation of the proposed PG discretization that uses a new bilinear form defined on *standard linear finite element spaces*. The new

formulation leads to strong connections with the upwinding FD methods and to a better understanding of both the FD and the FE methods.

The standard variational formulation for solving (1) with $\kappa = 1$ is: Find $u \in Q = H_0^1(0,1)$ such that

$$b(v,u) = \varepsilon\, a_0(u,v) + (u',v) = (f,v) \qquad \text{for all } v \in V = H_0^1(0,1). \tag{22}$$

A Petrov–Galerkin method for solving equation (22) chooses a test space $V_h \subset V = H_0^1(0,1)$ that, in general, could be different from the trial space $\mathcal{M}_h \subset Q = H_0^1(0,1)$.

4.1. **General bubble upwinding Petrov–Galerkin method**

On $[0,h]$, consider a continuous bubble generating function $B : [0,h] \to \mathbb{R}$ with the following properties:

$$B(0) = B(h) = 0, \tag{23}$$

$$\int_0^h B(x)\,dx = b_1 h \text{ with } b_1 > 0. \tag{24}$$

By translating B, we generate n bubble functions that are locally supported. For $i = 1, 2, \cdots, n$, we define $B_i : [0,1] \to \mathbb{R}$ by $B_i(x) = B(x - x_{i-1}) = B(x - (i-1)h)$ on $[x_{i-1}, x_i]$, and we extend it by zero to the entire interval $[0,1]$. Note that $B_1 = B$ on $[0,h]$, and for $i = 1, 2, \cdots, n$, we have

$$B_i(x_{i-1}) = B_i(x_i) = 0, \text{ and } B_i = 0 \text{ on } [0,1]\backslash(x_{i-1}, x_i). \tag{25}$$

In addition,

$$\int_{x_{i-1}}^{x_i} B_i(x)\,dx = b_1 h \text{ with } b_1 > 0. \tag{26}$$

Next, we consider a particular class of Petrov–Galerkin discretizations of the model problem (22) with trial space $\mathcal{M}_h = \text{span}\{\varphi_j\}_{j=1}^{n-1}$ and the test space V_h obtained by modifying $\mathcal{M}_h$ such that diffusion is created from the convection term.

The FE *bubble upwinding* idea is based on building V_h by translating a general function B satisfying (23) and (24). To be more precise, we define the test space V_h by

$$V_h := \text{span}\{\varphi_j + (B_j - B_{j+1}) \mid j = 1, 2, \cdots n-1\},$$

where $\{B_i\}_{i=1,\cdots,n}$ are defined above and satisfy (25) and (26). We note that both $\mathcal{M}_h$ and V_h have the same dimension of $(n-1)$.

The upwinding Petrov–Galerkin discretization with general bubble functions for (1) is: Find $u_h \in \mathcal{M}_h$ such that

$$b(v_h, u_h) = \varepsilon\, a_0(u_h, v_h) + (u_h', v_h) = (f, v_h) \qquad \text{for all } v_h \in V_h. \tag{27}$$

Next, we show that the variatonal formulation (27) admits a reformulation that uses a new bilinear form defined on *standard linear finite element spaces*. We look for

$$u_h = \sum_{j=1}^{n-1} \alpha_j \varphi_j,$$

and consider a generic test function

$$v_h = \sum_{i=1}^{n-1} \beta_i \varphi_i + \sum_{i=1}^{n-1} \beta_i (B_i - B_{i+1}) = \sum_{i=1}^{n-1} \beta_i \varphi_i + \sum_{i=1}^{n} (\beta_i - \beta_{i-1}) B_i,$$

where, we define $\beta_0 = \beta_n = 0$. By introducing the notation

$$B_h := \sum_{i=1}^{n} (\beta_i - \beta_{i-1}) B_i, \text{ and } w_h := \sum_{i=1}^{n-1} \beta_i \varphi_i,$$

we get

$$v_h = w_h + B_h.$$

By using the formulas (25), (26), and the facts that u_h', w_h' are constant on each of the intervals $[x_{i-1}, x_i]$, and that $w_h' = \frac{\beta_i - \beta_{i-1}}{h}$ on $[x_{i-1}, x_i]$, we obtain

$$(u_h', B_h) = \sum_{i=1}^{n} \int_{x_{i-1}}^{x_i} u_h' (\beta_i - \beta_{i-1}) B_i = \sum_{i=1}^{n} u_h'\, w_h'\, h \int_{x_{i-1}}^{x_i} B_i = b_1 h \sum_{i=1}^{n} \int_{x_{i-1}}^{x_i} u_h' w_h'.$$

Thus,

$$(u_h', B_h) = b_1 h (u_h', w_h'), \text{ where } v_h = w_h + B_h. \tag{28}$$

In addition, since u_h' is constant on $[x_{i-1}, x_i]$, we have

$$(u_h', B_i') = u_h' \int_{x_{i-1}}^{x_i} B_i'(x)\, dx = \ 0 \text{ for all } i = 1, 2, \cdots, n.$$

Hence,

$$(u_h', B_h') = 0, \text{ for all } u_h \in \mathcal{M}_h, v_h = w_h + B_h \in V_h. \tag{29}$$

From (28) and (29), for any $u_h \in \mathcal{M}_h$ and $v_h = w_h + B_h \in V_h$, we get

$$b(v_h, u_h) = (\varepsilon + b_1 h)\, (u_h', w_h') + (u_h', w_h). \tag{30}$$

The addition of the bubble part to the test space leads to the extra diffusion term $b_1 h (u_h', w_h')$ with $b_1 h > 0$ matching the sign of the coefficient of u' in (1). This justifies the terminology of *upwinding PG* method.

Here are two important notes regarding the variational formulation (30). First, note that only the linear part w_h of v_h appears in the expression of $b(v_h, u_h)$ of (30). Second, note that the functional $v_h \to (f, v_h)$ can be also viewed as a functional only of the linear part w_h. Indeed, using the splitting $v_h = w_h + B_h$ with $B_h := \sum_{i=1}^{n}(\beta_i - \beta_{i-1})B_i$, we have

$$(f, v_h) = (f, w_h) + \left(f, \sum_{i=1}^{n} h w_h' B_i\right) = (f, w_h) + h\left(f, w_h' \sum_{i=1}^{n} B_i\right).$$

Consequently, the variational formulation of the upwinding Petrov–Galerkin method can be reformulated as: Find $u_h \in \mathcal{M}_h$ such that

$$(31) \quad (\varepsilon + b_1 h)(u_h', w_h') + (u_h', w_h) = (f, w_h) + h\left(f, w_h' \sum_{i=1}^{n} B_i\right), w_h \in \mathcal{M}_h.$$

Remark 4.1. The reformulation (31) of the upwinding PG discretization (27) involves the same piecewise linear test and trial space.The coefficient of (u_h', w_h'), (that we call diffusion coefficient) in (31) is $d = d_{\varepsilon,h} = \varepsilon + h\, b_1$. Thus, the left-hand side of (31) is given by the bilinear form $b_d(u_h, w_h)$ defined only for continuous piecewise linear functions. Consequently, for the given test space V_h, the optimal trial norm on $\mathcal{M}_h$ is

$$(32) \qquad \|u_h\|_{*,h}^2 = (\varepsilon + h\, b_1)^2 |u_h|^2 + |u_h|_{*,h}^2,$$

where $|u_h|_{*,h}^2$ is defined in (16), see Section 3.2.

The reformulation (31) leads to the linear system

$$(33) \qquad \left(\left(\frac{\varepsilon}{h} + b_1\right)S + C\right) U = F_{PG},$$

where $U, F_{PG} \in \mathbb{R}^{n-1}$ with:

$$U := \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_{n-1} \end{bmatrix}, \quad F_{PG} := \begin{bmatrix} (f, \varphi_1) \\ (f, \varphi_2) \\ \vdots \\ (f, \varphi_{n-1}) \end{bmatrix} + \begin{bmatrix} (f, B_1 - B_2) \\ (f, B_2 - B_3) \\ \vdots \\ (f, B_{n-1} - B_n) \end{bmatrix},$$

and S, C are the matrices defined in (8), for the FD discretization.

Remark 4.2. Here, we note that by using the notation

$$d_{\varepsilon,h} = \varepsilon + h\, b_1, \text{ or } \frac{\varepsilon}{h} + b_1 = \frac{d_{\varepsilon,h}}{h},$$

the matrix of the finite element system (33) is

$$(34) \qquad M_{fe} = \text{tridiag}\Big(-\frac{d_{\varepsilon,h}}{h} - \frac{1}{2},\ \frac{2\, d_{\varepsilon,h}}{h},\ -\frac{d_{\varepsilon,h}}{h} + \frac{1}{2}\Big).$$

4.2. **Comparing the upwinding FD and the PG FE methods**

In this section, we compare the general upwinding Petrov–Galerkin finite element discretization (27) with the general upwinding finite difference method (6) with $\varepsilon_h = \varepsilon(1+\Phi(\mathbb{P}e))$ as defined in Section 2. We consider that the bubble B for the PG method is chosen such that $\varepsilon_h = \varepsilon + h\,b_1$, i.e.,

$$\Phi(\mathbb{P}e) = \Phi\Big(\frac{h}{2\varepsilon}\Big) = 2\,b_1\,\mathbb{P}e,$$

where b_1 is defined by (24). In this case, $\varepsilon_h = d_{\varepsilon,h}$, and a direct consequence of the matrix formulas (9) and (34), is that the FD matrix of the system (7) coincides with the FE matrix of the system (33), i.e.,

$$M_{fd} = M_{fe}. \tag{35}$$

Since b_1 is allowed to depend on ε and h, the function $\Phi(\mathbb{P}e) = 2\,b_1\,\mathbb{P}e$ can recover the functions Φ used for the classical upwinding finite difference methods. Furthermore, the *artificial diffusion*, $\varepsilon\Phi(\mathbb{P}e)$ that is introduced for upwinding FD discretization, becomes exactly the integral of a bubble function that defines the corresponding upwinding PG method, i.e.,

$$\varepsilon\Phi(\mathbb{P}e) = h\,b_1 = \int_0^h B.$$

Next, we show that the FD system (7) and the FE system (33) correspond to variational formulations that use the same bilinear form defined on $\mathcal{M}_h\times\mathcal{M}_h$ and compare the right-hand side functionals.

For a continuous function $\theta : [0,1]\to\mathbb{R}$ such that $\theta(0)=\theta(1)=0$, the composite trapezoid rule (CTR) on the uniformly distributed nodes $x_0, x_1, \ldots, x_n$ is

$$\int_0^1 \theta(x)dx \approx T_n(\theta) := h\sum_{i=1}^{n-1}\theta(x_i). \tag{36}$$

If the vector $u^{FD} = [u_1, \ldots, u_{n-1}]^T$ is the solution of the finite difference system (6), then we define the corresponding proxy function $u_h^{FD} \in \mathcal{M}_h$ by

$$u_h^{FD} := \sum_{i=1}^{n-1} u_i\varphi_i.$$

From Section 2, we have that u^{FD} is the solution of the system

$$M_{fd}\,u^{FD} = h\,F, \text{ where } F = [f(x_1), \cdots, f(x_{n-1})]^T. \tag{37}$$

Elementary calculations show that

$$h\,f(x_j) = T_n(f\varphi_j) = T_n(f(\varphi_j + B_j - B_{j+1})), \text{ for } j = 1, 2, \cdots, n-1. \tag{38}$$

The following remark illustrates another connection between the upwinding FD and the bubble PG FE method.

Remark 4.3. In light of (35) and (38), we note that the upwinding FD system (6) can be obtained from the PG discretization (27) reformulated as (31) and leading to the system (33), where the entries of the RHS vector, $\int_0^1 f(\varphi_j + B_j - B_{j+1})$ are approximated with the CTR approximations.

At the local level, this corresponds to using the standard trapezoid rule to approximate each of the integrals

$$\int_{x_{j-1}}^{x_j} f\,\varphi_j, \quad \int_{x_j}^{x_{j+1}} f\,\varphi_j, \text{ and } \int_{x_{j-1}}^{x_j} f\,B_j. \tag{39}$$

Next, we will justify that, under the assumption $\Phi(\mathbb{P}e) = 2\,b_1\,\mathbb{P}e$, the upwinding FD method (6) and the PG discretization (27) can be viewed as variational formulations using the same bilinear form defined on $\mathcal{M}_h \times \mathcal{M}_h$. First, for the FD formulation, using equation (38), the algebra of Section 3.1, with $\varepsilon \to \varepsilon_h$, and the notation introduced in Section 3.2, we have that the system (37) of the upwinding FD method (6) corresponds to the variational formulation

$$b_d(u_h^{FD}, \varphi_j) = T_n(f\,\varphi_j) = T_n(f(\varphi_j + B_j - B_{j+1})), \quad \text{for all } j = 1, 2, \ldots, n-1,$$

with $d = \varepsilon_h = d_{\varepsilon,h} = \varepsilon + h\,b_1$. Using the linearity of $b_d(u_h^{FD}, \cdot)$ and $T_n(f\,\cdot)$, we can further conclude that

$$b_d(u_h^{FD}, w_h) = T_n(f\,w_h), \text{ for all } w_h \in \mathcal{M}_h. \tag{40}$$

Second, for the FE formulation, if $u_h = u_h^{FE}$ is the solution of (27), using the equivalent variational formulation (31) with $d = d_{\varepsilon,h} = \varepsilon + h\,b_1$, we have

$$b_d(u_h^{FE}, w_h) = (f, w_h) + h\left(f, w_h' \sum_{i=1}^{n} B_i\right), \text{ for all } w_h \in \mathcal{M}_h. \tag{41}$$

Based on the reformulations (40) and (41), we can estimate now the difference between the upwinding FD and the bubble PG solutions in the optimal trial norm. We define the linear functionals $F_1, F_2 : \mathcal{M}_h \to \mathbb{R}$ by

$$F_2(w_h) = (f\,w_h) + h\left(f, w_h' \sum_{i=1}^{n} B_i\right), \text{ and}$$

$$F_1(w_h) = T_n\left(f\,w_h + h\,f\,w_h' \sum_{i=1}^{n} B_i\right) = T_n(f\,w_h).$$

The functionals $F_h, W_h : \mathcal{M}_h \to \mathbb{R}$ defined by

$$F_h(w_h) = \int_0^1 f\,w_h\,dx - T_n(f\,w_h), \text{ and } W_h(w_h) = h\left(f, w_h' \sum_{i=1}^{n} B_i\right)$$

are also linear functionals and

$$F_2(w_h) - F_1(w_h) = F_h(w_h) + W_h(w_h).$$

Using Remark 3.1, and the norm $\|\cdot\|_{*,h}$ described in (32), we obtain

$$\begin{aligned} \|u_h^{FE} - u_h^{FD}\|_{*,h} &= \|F_2 - F_1\|_{\mathcal{M}_h^*} = \|F_h + W_h\|_{\mathcal{M}_h^*} \\ &\leq \|F_h\|_{\mathcal{M}_h^*} + \|W_h\|_{\mathcal{M}_h^*}. \end{aligned} \tag{42}$$

It was proved in [7], by using standard approximation properties of Trapezoid Rule, that if $f \in \mathcal{C}^2([0,1])$, then

$$\|F_h\|_{\mathcal{M}_h^*} \leq h^2 \left(\frac{\|f''\|_\infty}{12} + \frac{\|f'\|_\infty}{6} \right). \tag{43}$$

On the other hand, using the Cauchy–Schwartz inequality and assuming that $\|B\|_\infty \leq M$, it is easy to check that

$$\|W_h\|_{\mathcal{M}_h^*} \leq M\, h \|f\|_{L^2}.$$

Consequently, we obtain

$$\|u_h^{FE} - u_h^{FD}\|_{*,h} \leq h^2 \left(\frac{\|f''\|_\infty}{12} + \frac{\|f'\|_\infty}{6} \right) + M\, h \|f\|_{L^2}. \tag{44}$$

The next remark provides a way to improve a standard upwinding FD method by using its connections with a bubble PG method.

Remark 4.4. The estimate (44) is suboptimal in the sense that the second term in the RHS of (44) is only $O(h)$. In light of Remark 4.3, this can be improved by using a modified upwinding FD method as follows. We start with the corresponding PG discretization (27), reformulated as (31), where the bubble B is chosen such that $\varepsilon_h = \varepsilon + h\, b_1$ (or $\Phi(\mathbb{P}e) = 2\, b_1 \mathbb{P}e$), leading to the system (33). Then the RHS entries of (33), i.e., $\int_0^1 f(\varphi_j + B_j - B_{j+1})$, are approximated with better (than trapezoid) quadratures. In fact, this suggests to use a higher order quadrature, (e.g., Cavalieri–Simpson or Gaussian quadrature) to approximate each of the integrals in (39). In this way, new upwinding FD methods can be obtained by better approximating the dual vector of upwinding PG methods. Besides an improvement of the order of the estimate (44), numerical computations show that the bubble PG approximation is better than related upwinding FD methods even in the discrete infinity error.

Consequently, the upwinding FD method (6) for solving (1) can be improved just by modifying the RHS vector of the FD system (7). The new j-th entry of the RHS system (7) is obtained by approximating the integral of $(f, \varphi_j + B_j - B_{j+1})$ using, locally, higher order quadratures. Since we can choose

different bubble functions B and different quadratures to locally approximate the dual vectors, the improvement process is not unique.

On the other hand, any bubble PG method uses a fixed quadrature to approximate the dual vector. Thus, numerically, this performs identically with an upwinding FD method with a special RHS induced by the quadrature.

5. UPWINDING PG WITH QUADRATIC BUBBLE FUNCTIONS

In this section, we consider the model problem (10) with the discrete space $\mathcal{M}_h = span\{\varphi_j\}_{j=1}^{n-1}$ and V_h a modification of $\mathcal{M}_h$ using *quadratic bubble functions*. The resulting method can be found in e.g., [3, 14, 23, 20, 29]. However, based on the findings of the previous section, we relate the *quadratic bubble PG* method to the general upwinding FD method and present ways to improve the performance of upwinding FD method.

First, for a parameter $\beta > 0$, we define the bubble function B on $[0, h]$ by

$$B(x) = \frac{4\,\beta}{h^2} x(h - x).$$

Elementary calculations show that (24) holds with $b_1 = \frac{2\beta}{3}$. Using the function B and the general construction of Section 4.1, we define the set of bubble functions $\{B_1, B_2, \cdots, B_n\}$ on $[0, 1]$ and

$$V_h := \text{span}\{\varphi_j + (B_j - B_{j+1})\}_{j=1}^{n-1}.$$

In this case, we have $d_{\varepsilon,h} = \varepsilon + \frac{2\beta}{3}h$. According to (34), we obtain

$$M_{fe} = \text{tridiag}\left(-\frac{\varepsilon}{h} - \frac{2\beta}{3} - \frac{1}{2},\ \frac{2\,\varepsilon}{h} + \frac{4\beta}{3},\ -\frac{\varepsilon}{h} - \frac{2\beta}{3} + \frac{1}{2}\right).$$

For $\varepsilon_h = d_{\varepsilon,h} = \varepsilon + \frac{2\beta}{3}h$, or $\Phi(\mathbb{P}e) = 2\,b_1\mathbb{P}e = \frac{4\beta}{3}\mathbb{P}e$, as presented at the beginning of Section 4.2, the matrix of the system (7) is $M_{fd} = M_{fe}$.

Here, we note that we can relate any upwinding FD method defined by an admissible function $\Phi(\cdot)$ to our quadratic bubble PG method introduced in this section. This is justified by the fact that we can choose β, hence $b_1 = \frac{2\beta}{3}$, such that $\varepsilon_h = \varepsilon(1 + \Phi(\mathbb{P}e)) = \varepsilon + b_1\,h$, i.e.,

$$\beta = \frac{3}{4}\frac{\Phi(\mathbb{P}e)}{\mathbb{P}e}.$$

Now, we can benefit from Remark 4.4. By using, for example, the Cavalieri–Simpson (CS) rule, we can improve the upwinding FD method (6) for solving (1) with $\varepsilon_h = \varepsilon + \frac{2\beta}{3}h$ by modifying only the RHS vector of the FD system (7).

Next, the new j-th entry of the RHS system (7) is obtained by approximating $(f, \varphi_j + B_j - B_{j+1})$ using the CS rule

$$\int_a^b g(x)\,dx \approx \frac{b-a}{6}\left(g(a) + 4g\left(\frac{a+b}{2}\right) + g(b)\right),$$

on each mesh interval. This leads to replacing $hf(x_j)$ in (7) by the value

$$\frac{h}{3}\big[(1+2\,\beta)f(x_j - h/2) + f(x_j) + (1-2\,\beta)f(x_j + h/2)\big].$$

As a specific application, we consider the case $\beta = 3/4$ that leads to $d_{\varepsilon,h} = \varepsilon_h = \varepsilon + h/2$. In this case, the FE matrix of the system (33) coincides with the matrix of the standard upwinding FD system (5), and we have

$$M_{fe} = M_{fd} = \text{tridiag}\left(-\frac{\varepsilon}{h} - 1,\ \frac{2\,\varepsilon}{h} + 1,\ -\frac{\varepsilon}{h}\right).$$

To improve the performance of the upwinding FD method (5), we can consider the *CS quadratic upwinding method* that solves the system

$$M_{fd}\,U = G, \text{ where}$$

$$G_j = \frac{h}{3}\left(\frac{5}{2}f(x_j - h/2) + f(x_j) - \frac{1}{2}f(x_j + h/2)\right),\ j = 1, 2, \cdots, n-1.$$

Numerical results show that the *Cavalieri–Simpson FD (CS-FD) method* performs better than the standard upwinding FD method even if we measure the error in the discrete infinity norm. We note that the standard upwinding FD method is in fact the Trapezoid Finite Difference (T-FD) as a result of the quadratic bubble PG method using the Trapezoid rule for estimating the dual vector. We solved (1) with $\kappa = 1$ and $f(x) = 2x$, which allows to find the exact solution and to compute the discrete infinity error approximation. For example, for $\epsilon = 10^{-6}$, the standard upwinding FD method, or T-FD produces a discrete infinity error of order $O(h)$, while the CS-FD method exhibits higher order using the same discrete infinity error. For example, when $n = 800$, the discrete infinity error for T-FD is 0.0012, and for CS-FD, the error is 0.64×10^{-6}. For the two methods, the error behaviour and error order, in various norms, will be addressed in future work.

6. UPWINDING PG WITH EXPONENTIAL BUBBLE FUNCTIONS

We consider the model problem (10) with the discrete space $\mathcal{M}_h = \text{span}\{\varphi_j\}_{j=1}^{n-1}$ and V_h a modification of $\mathcal{M}_h$ by using an *exponential bubble function*. We define the bubble function B on $[0, h]$ as the solution of

$$-\varepsilon B'' - B' = 1/h,\ B(0) = B(h) = 0. \tag{45}$$

Using the function B and the general construction of Section 4.1, we define the set of bubble functions $\{B_1, B_2, \cdots, B_n\}$ on $[0,1]$ and

$$V_h := \text{span}\{\varphi_j + (B_j - B_{j+1})\}_{j=1}^{n-1} = \text{span}\{g_j\}_{j=1}^{n-1}, \tag{46}$$

where $g_j := \varphi_j + (B_j - B_{j+1})$, $j = 1, 2, \cdots, n-1$.

In order to address efficient computations of coefficients and the finite element matrix of the *exponential bubble PG method*, we introduce the following notation

$$g_0 := \tanh(\mathbb{P}e) = \frac{e^{\frac{h}{2\varepsilon}} - e^{-\frac{h}{2\varepsilon}}}{e^{\frac{h}{2\varepsilon}} + e^{-\frac{h}{2\varepsilon}}} = \frac{1 - e^{-\frac{h}{\varepsilon}}}{1 + e^{-\frac{h}{\varepsilon}}}, \tag{47}$$

$$l_d := \frac{1+g_0}{2} = \frac{e^{\frac{h}{\varepsilon}} - 1}{e^{\frac{h}{\varepsilon}} + e^{-\frac{h}{\varepsilon}}}, \text{ and } l_0 := \frac{1+g_0}{2g_0} = \frac{l_d}{g_0}, \tag{48}$$

$$u_d := \frac{1-g_0}{2} = \frac{1 - e^{-\frac{h}{\varepsilon}}}{e^{\frac{h}{\varepsilon}} + e^{-\frac{h}{\varepsilon}}}, \text{ and } u_0 := \frac{1-g_0}{2g_0} = \frac{u_d}{g_0}. \tag{49}$$

It is easy to check that the unique solution of (45) is

$$B(x) = l_0 \left(1 - e^{-\frac{x}{\varepsilon}}\right) - \frac{x}{h}, \ x \in [0,h], \text{ and} \tag{50}$$

$$\int_0^h B(x)\, dx = \frac{h}{2g_0} - \varepsilon. \tag{51}$$

Consequently, we have that (24) holds with $b_1 = \frac{1}{2g_0} - \frac{\varepsilon}{h}$, and we obtain that $\varepsilon + b_1\, h = \frac{h}{2g_0}$. Using Remark 4.1, the optimal trial norm on $\mathcal{M}_h$ is

$$\|u_h\|_{*,h}^2 = \left(\frac{h}{2g_0}\right)^2 |u_h|^2 + |u_h|_{*,h}^2 \tag{52}$$

where $|u_h|_{*,h}^2$ is defined in (16).

Using Remark 4.2, we obtain that $\frac{\varepsilon_{b,h}}{h} = \frac{1}{2g_0}$, and the matrix for the PG finite element discretization with exponential bubble test space becomes

$$\begin{aligned} M_{fe}^e &= \text{tridiag}\left(-\frac{1+g_0}{2g_0},\ \frac{1}{g_0},\ -\frac{1-g_0}{2g_0}\right) \\ &= \text{tridiag}\left(-l_0, 1/g_0, -u_0\right) = \frac{1}{g_0}\ \text{tridiag}\left(-l_d, 1, -u_d\right). \end{aligned} \tag{53}$$

For $\varepsilon_h = d_{\varepsilon,h} = \varepsilon + b_1 h = \dfrac{h}{g_0}$, we have $\Phi(\mathbb{P}e) = 2\, b_1 \mathbb{P}e = \mathbb{P}e\, \coth(\mathbb{P}e) - 1$. According to Section 4.2, the matrix of the system (7) is $M_{fd}^e = M_{fe}^e$.

By applying Remark 4.3, using the trapezoid rule to approximate the dual vector for the PG method, we get the upwinding FD method known as the *Il'in–Allen–Southwell (IAS) method*, according to [28], or to the *Scharfetter–Gummel (SG) method*, according to [26].

Using again Remark 4.4, and the Cavalieri–Simpson rule, we can improve the upwinding FD method (6) for solving (1) with $\varepsilon_h = \frac{h}{g_0}$ (which is the IAS or the SG method), by modifying only the RHS vector of the FD system (7). The new j-th entry of the RHS system (7) is obtained by approximating $(f, \varphi_j + B_j - B_{j+1})$ using the Cavalieri–Simpson rule, on each mesh interval. Since the B_j functions are generated by the exponential bubble function B given by (50), this leads to replacing $hf(x_j)$ in (7) for $j = 1, 2, \cdots, n-1$, with

$$G_j := \frac{h}{3}\left[(1 + 2\,B(h/2))f(x_j - h/2) + f(x_j) + (1 - 2\,B(h/2))f(x_j + h/2)\right].$$

The *CS exponential upwinding method* reduces to solving for U the system

$$M_{fd}^{e}\, U = G.$$

As in the polynomial bubble case, the CS-FD method performs better than the standard IAS or SG method. It is important to mention here that the upwinding PG method based on the exponential bubble produces in fact the exact solution at the nodes, provided that the dual vector is computed exactly. Variants of this result seem to be known in various forms, see e.g., [28, 27]. We include a simple proof of the above statement that is based on the properties of the exponential bubble function B. In order to proceed with the proof, we will need to emphasize a few properties of the test functions g_j as follows. For any $j = 1, 2, \cdots, n$, we have

$$g_j = \begin{cases} B_j + \varphi_j & \text{if } x \in [x_{j-1}, x_j], \\ -B_{j+1} + \varphi_j & \text{if } x \in [x_j, x_{j+1}], \end{cases} \quad \text{and} \tag{54}$$

(55)

$$g_j^{'} = \begin{cases} B_j^{'} + \dfrac{1}{h} & \text{if } x \in (x_{j-1}, x_j), \\ -B_{j+1}^{'} - \dfrac{1}{h} & \text{if } x \in (x_j, x_{j+1}). \end{cases} \qquad g_j^{''} = \begin{cases} B_j^{''} & \text{if } x \in (x_{j-1}, x_j), \\ -B_{j+1}^{''} & \text{if } x \in (x_j, x_{j+1}). \end{cases}$$

Using (55) and the fact that on $[x_{j-1}, x_j]$ the functions B_j satisfy the same differential equations as B, see (45), we obtain

$$-\varepsilon g_j'' - g_j' = 0, \text{ on } (x_{j-1}, x_j) \cup (x_j, x_{j+1}), \tag{56}$$

$$g_j'(x_j-) - g_j'(x_j+) = B'(h) + B'(0) - \frac{2}{h} = \frac{1}{\varepsilon\, g_0}, \tag{57}$$

$$g_j'(x_{j-1}) = g_j'(x_{j-1}+) = B'(0) + \frac{1}{h} = \frac{1}{\varepsilon}\,\frac{l_d}{g_0}, \text{ and} \tag{58}$$

$$g_j'(x_{j+1}) = g_j'(x_{j+1}-) = -B'(h) - \frac{1}{h} = -\frac{1}{\varepsilon}\,\frac{u_d}{g_0}. \tag{59}$$

Next, we are ready to prove the following result.

THEOREM 6.1. *Let* $u_h := \sum_{i=1}^{n-1} u_i\varphi_i$ *be the finite element solution of* (27) *with the test space as defined in* (46). *Then,* u_h *coincides with the linear interpolant* $I_h(u)$ *of the exact solution* u *of* (1), *with* $\kappa = 1$ *on the nodes* $x_0, x_1, \ldots, x_n$. *In other words,* $u_j = u(x_j)$, $j = 1, 2, \cdots, n-1$.

Proof. For any fixed $j \in \{1, 2, \cdots, n-1\}$, we multiply the differential equation (1) (with $\kappa = 1$) by g_j and integrate by parts on the interval $[x_{j-1}, x_{j+1}]$ to obtain

$$\varepsilon \int_{x_{j-1}}^{x_{j+1}} u' g_j' - \int_{x_{j-1}}^{x_{j+1}} u g_j' + (u g_j)|_{x_{j-1}}^{x_{j+1}} = \int_{x_{j-1}}^{x_{j+1}} f g_j = (f, g_j). \tag{60}$$

Using that $g_j(x_{j-1}) = g_j(x_{j+1}) = 0$, the third term in the LHS of (60) is zero. Next, we apply integration by parts for both integrals in the LHS of (60), splitting the integration on two subintervals $[x_{j-1}, x_j]$ and $[x_j, x_{j+1}]$, such that g_j is smooth enough, to obtain

$$\begin{aligned} &\int_{x_{j-1}}^{x_j} (-\varepsilon g_j'' - g_j')\, u + \int_{x_j}^{x_{j+1}} (-\varepsilon g_j'' - g_j')\, u \\ &+ \varepsilon\,(u\, g_j')|_{x_{j-1}}^{x_j} + \varepsilon\,(u\, g_j')|_{x_j}^{x_{j+1}} = (f, g_j). \end{aligned} \tag{61}$$

By using (56), from (61) we get

$$\begin{aligned} &-\varepsilon g_j'(x_{j-1})\, u(x_{j-1}) + \varepsilon[g_j'(x_j-) - g_j'(x_j+))\, u(x_j] \\ &-\varepsilon g_j'(x_{j+1})\, u(x_{j+1}) = (f, g_j). \end{aligned} \tag{62}$$

Combining (62) with (57)-(59), we obtain

$$-\frac{l_d}{g_0}\, u(x_{j-1}) + \frac{1}{g_0}\, u(x_j) - \frac{u_d}{g_0}\, u(x_{j+1}) = (f, g_j),\ j = 1, \cdots, n-1. \tag{63}$$

Here, we notice that the matrix of the system (63) with the vector unknown $U_e = [u(x_1), \cdots u(x_{n-1})]^T$, coincides with the matrix M_{fe}^e of the system solving for the finite element solution of (27) with exponential bubble test space (see (53)). Since the right-hand sides of the two systems are the same and equal to $[(f, g_1), \cdots, (f, g_{n-1})]^T$, and M_{fe}^e is invertible, we conclude that $u_j = u(x_j)$, $j = 1, 2, \cdots, n-1$. $\square$

As a consequence of Theorem 6.1, using Remark 4.4 and a similar technique used for obtaining (43), we state the following result.

THEOREM 6.2. *Let $f \in \mathcal{C}^{(m+1)}([0,1])$ and assume that an upwinding FD method is obtained from the exponential PG FE method by using a quadrature of order $O(h^{m+1})$, on each mesh interval, to approximate the dual vector $[(f,g_1),\cdots,(f,g_{n-1})]^T$. Then*

$$\|I_h(u) - u_h^{FD}\|_{*,h} \leq C\, h^m, \tag{64}$$

where C is a constant that depends on f, B, and their derivatives.

We note that $|v_h(x)| \leq |v_h|$ for all $x \in [0,1]$ and all $v_h \in \mathcal{M}_h$, and from the representation (52) of the discrete optimal norm $\|\cdot\|_{*,h}$, we have

$$\|v_h\|_{\infty,h} := \max_{i=\overline{0,n}} |v_h(x_i)| \leq |v_h| \leq \frac{2g_0}{h}\|v_h\|_{*,h}, \text{ for all } v_h \in \mathcal{M}_h.$$

As a consequence of the Theorem 6.2 and the above estimate, we obtain the following result.

COROLLARY 6.3. *Under the assumptions of Theorem 6.2, we have*

$$\|I_h(u) - u_h^{FD}\|_{\infty,h} \leq 2g_0\, C\, h^{m-1}, \tag{65}$$

where C is the constant used for (64).

Numerical tests show that the estimate (65) does not hold if $\varepsilon << h$, and u_h^{FD} is replaced by the computed solution $u_{h,c}^{FD}$. This can be justified based on the error in computing $e^{-\frac{h}{\varepsilon}}$. We note that if $\frac{h}{\varepsilon}$ is too large, then $e^{-\frac{h}{\varepsilon}}$ is computed as 0. For example, for the double precision arithmetic, we have that $e^{-36.05}$ is smaller than the ϵ-machine. Thus, $1 + e^{-\frac{h}{\varepsilon}}$ is computed as 1 for $\frac{h}{\varepsilon} \geq 36.05$. Using standard calculus limits, we have that for $\frac{\varepsilon}{h} \to 0$,

$$g_0 \to 1, \text{ and } g_j = \varphi_j + B_j - B_{j+1} \to \chi_{|_{[x_{j-1},x_j]}}. \text{ Consequently, for } \frac{\varepsilon}{h} \to 0$$

$$M_{fe}^e \to \text{tridiag}(-1,1,0), \text{ and } (f,g_j) \to \int_{x_{j-1}}^{x_j} f(x)\,dx.$$

Based on our observations, the computed matrix M_{fe}^e becomes tridiag$(-1,1,0)$, if $\varepsilon << h$. Using a high order quadrature to estimate the dual vector of the exponential bubble PG method leading to an upwinding FD method, we can get a very accurate approximation of $(f,g_j) \approx \int_{x_{j-1}}^{x_j} f(x)\,dx$, especially if f is, for example, a polynomial function. Thus, the computed linear system is very close or identical to the system

$$[\text{tridiag}(-1,1,0)]\,U = \left[\int_{x_0}^{x_1} f(x)\,dx, \cdots, \int_{x_{n-2}}^{x_{n-1}} f(x)\,dx\right]^T.$$

The system can be solved exactly to obtain

$$u_j = \int_0^{x_j} f(x)\,dx,\ j = 1, 2, \cdots, n-1.$$

This implies that, when $\varepsilon << h$, the component u_j of the computed PG discrete solution is very close or identical to the value $w(x_j)$ where $w(x) = \int_0^x f(t)\,dt$. By decreasing h, as long as $\frac{\varepsilon}{h}$ is still very small, the computed discrete solution remains close to the interpolant of w on $[0, x_{n-1}]$. Thus, by taking the discrete infinity norm on the nodes $x_1, \cdots, x_{n-1}$, we have

$$\|I_h(u) - u_{h,c}^{FD}\|_{\infty,h} \approx \|I_h(u) - I_h(w)\|_{\infty,h} = \|I_h(u-w)\|_{\infty,h},$$

and the difference $\|I_h(u-w)\|_{\infty,h}$ approaches $\|(u-w)_{|_{[x_1,x_{n-1}]}}\|_\infty$ for $h \to 0$.

Consequently, the error $\|I_h(u) - u_{h,c}^{FD}\|_{\infty,h}$ is less sensitive to changes in $h \to 0$, as long as $\frac{\varepsilon}{h}$ is very small, and (65) cannot be checked numerically.

We discretized (1) for $\kappa = 1$, $f(x) = 2x$, $\epsilon = 10^{-6}$, and for the exponential bubble PG method. We used the Gaussian quadrature G_3, with three nodes to locally approximate the dual vector. For the following values of h, $h = \frac{1}{100}, \frac{1}{200}, \frac{1}{400}, \frac{1}{800}, \frac{1}{1600}$, we obtained

$$\|I_h(u) - u_{h,c}^{FD}\|_{\infty,h} \approx 2 \times 10^{-6}.$$

In conclusion, if the upwinding PG with exponential bubble, or its FD versions are implemented, then we should avoid choosing $h >> \varepsilon$. For a given $\varepsilon << 1$, in order to expect decreasing discrete infinity error, we can choose for example $h \leq 30\varepsilon$.

REFERENCES

[1] C. Bacuta, D. Hayes, and J. Jacavage, *Notes on a saddle point reformulation of mixed variational problems*. Comput. Math. Appl. **95** (2021), 4–18.

[2] C. Bacuta, D. Hayes, and J. Jacavage, *Efficient discretization and preconditioning of the singularly perturbed reaction-diffusion problem*. Comput. Math. Appl. **109** (2022), 270–279.

[3] C. Bacuta, D. Hayes, and T. O'Grady, *Saddle point least squares discretization for convection-diffusion*. Appl. Anal. **103** (2024), *12*, 2241–2268.

[4] C. Bacuta and J. Jacavage, *A non-conforming saddle point least squares approach for an elliptic interface problem*. Comput. Methods Appl. Math. **19** (2019), *3*, 399–414.

[5] C. Bacuta and J. Jacavage, *Saddle point least squares preconditioning of mixed methods*. Comput. Math. Appl. **77** (2019), *5*, 1396–1407.

[6] C. Bacuta and K. Qirko, *A saddle point least squares approach for primal mixed formulations of second order PDEs*. Comput. Math. Appl. **73** (2017), *2*, 173–186.

[7] Cr. Bacuta and C. Bacuta, *Connections between finite difference and finite element approximations*. Appl. Anal. **102** (2023), *6*, 1808–1820.

[8] S. Bartels, *Numerical Approximation of Partial Differential Equations*. Texts Appl. Math. 64, Springer, Cham, 2016.

[9] D. Boffi, F. Brezzi, and M. Fortin, *Mixed Finite Element Methods and Applications*. Springer Ser. Comput. Math. 44, Springer, Heidelberg, 2013.

[10] D. Braess, *Finite Elements. Theory, Fast Solvers, and Applications in Solid Mechanics*. Cambridge Univ. Press, Cambridge, 1997.

[11] S. Brenner and L.R. Scott, *The Mathematical Theory of Finite Element Methods*. Texts Appl. Math. 15, Springer, New York, 1994.

[12] D. Broersen and R. Stevenson, *A robust Petrov–Galerkin discretisation of convection-diffusion equations*. Comput. Math. Appl. **68** (2014), *11*, 1605–1618.

[13] J. Chan, N. Heuer, T. Bui-Thanh, and L. Demkowicz, *A robust DPG method for convection-dominated diffusion problems II: adjoint boundary conditions and mesh-dependent test norms*. Comput. Math. Appl. **67** (2014), *4*, 771–795.

[14] I. Christie, D.F. Griffiths, A.R. Mitchell, and O.C. Zienkiewicz, *Finite element methods for second order differential equations with significant first derivatives*. Internat. J. Numer. Methods Engrg. **10** (1976), *6*, 1389–1396.

[15] A. Cohen, W. Dahmen, and G. Welper, *Adaptivity and variational stabilization for convection-diffusion equations*. ESAIM Math. Model. Numer. Anal. **46** (2012), *5*, 1247–1273.

[16] L.F. Demkowicz, *Mathematical Theory of Finite Elements*. Computational Science & Engineering, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2024.

[17] L. Demkowicz and J. Gopalakrishnan, *A class of discontinuous Petrov–Galerkin methods. Part I: the transport equation*. Comput. Methods Appl. Mech. Engrg. **199** (2010), *23-24*, 1558–1572.

[18] K. Eriksson, D. Estep, P. Hansbo, and C. Johnson, *Computational Differential Equations*. Cambridge Univ. Press, Cambridge, 1996.

[19] A. Ern and J-L. Guermond, *Theory and Practice of Finite Elements*. Appl. Math. Sci. 159, Springer, New York, 2004.

[20] R.H. Gallagher, O.C. Zienkiewicz and P. Hood, *Newtonian and non-newtonian viscous incompressible flow, temperature induced flows, finite element solutions, in the mathematics of finite elements and applications. II*. In : J.R. Whiteman (Ed.), *Proceedings of the Second Brunel University Conference of the Institute of Mathematics and its Applications* (Uxbridge, April 7–10, 1975), pp. 235–267. Academic Press, Inc., Harcourt Brace Jovanovich Publishers, London, New York, 1976.

[21] R. Lin and M. Stynes, *A balanced finite element method for singularly perturbed reaction-diffusion problems*. SIAM J. Numer. Anal. **50** (2012), *5*, 2729–2743.

[22] T. Linß, *Layer-adapted meshes for reaction-convection-diffusion problems*. Lecture Notes in Math. 1985, Springer, Berlin, 2010.

[23] A.R. Mitchell and D.F. Griffiths, *The Finite Difference Method in Partial Differential Equations*. A Wiley-Interscience Publication, John Wiley & Sons, Ltd., Chichester, 1980.

[24] K.W. Morton and J.W. Barrett, *Optimal Petrov–Galerkin methods through approximate symmetrization*. IMA J. Numer. Anal. **1** (1981), *4*, 439–468.

[25] J.T. Oden and L.F. Demkowicz, *Applied Functional Analysis*. Textb. Math., CRC Press, Boca Raton, FL, 2018.

[26] A. Quarteroni, R. Sacco and F. Saleri, *Numerical Mathematics*. Texts Appl. Math. 37, Springer, Berlin, 2007.

[27] H.G. Roos and M. Schopf, *Convergence and stability in balanced norms of finite element methods on Shishkin meshes for reaction-diffusion problems*. ZAMM Z. Angew. Math. Mech. **95** (2015), *6*, 551–565.

[28] H.G. Roos, M. Stynes, and L. Tobiska, *Numerical Methods for Singularly Perturbed Differential Equations*. Springer Ser. Comput. Math. 24, Springer, Berlin, 1996.

[29] O.C. Zienkiewicz, R.L. Taylor, and P. Nithiarasu, *The Finite Element Method for Fluid Dynamics*. Elsevier/Butterworth Heinemann, Amsterdam, 2014.

Cristina Bacuta
Constantin Bacuta
University of Delaware
Department of Mathematical Sciences
501 Ewing Hall
Newark, DE, USA, 19716
crbacuta@udel.edu
bacuta@udel.edu