Results in Applied Mathematics 8 (2020) 100105

Contents lists available at ScienceDirect

Results in Applied Mathematics

journal homepage: www.elsevier.com/locate/results-in-applied-mathematics

Saddle point least squares for the reaction–diffusion problem[☆]

Constantin Bacuta[a,*], Jacob Jacavage[b]

[a] University of Delaware, Mathematical Sciences, 501 Ewing Hall, Newark, DE 19716, United States of America
[b] Lafayette College, Department of Mathematics, Pardee Hall, Easton, PA 18042, United States of America

ARTICLE INFO

Article history:
Received 30 December 2019
Received in revised form 3 April 2020
Accepted 7 April 2020
Available online 25 April 2020

MSC:
74S05
74B05
65N22
65N55

Keywords:
Least squares
Saddle point systems
Mixed methods
Multilevel methods
Uzawa algorithm
Conjugate gradient
Singularly perturbed

ABSTRACT

We consider a mixed variational formulation for the reaction–diffusion problem based on a saddle point least square approach with an optimal test norm and nonconforming trial spaces. An Uzawa type iterative process for solving the discrete mixed formulations is proposed and choices for discrete stable spaces are provided. The implementation requires a nodal basis only for the test space, and assembly of a global saddle point system is avoided. For the test space, we use piecewise linear spaces of functions on Shishkin type meshes that provide almost optimal approximation in the standard symmetric elliptic formulation. Our saddle point least squares method has the advantage that the order of approximation of the solution in a balanced norm is improved if compared with the standard variational approach. Numerical results are included to support the proposed method.

1. Introduction

The Saddle Point Least Squares (SPLS) method and its versions was developed in [1–4]. In this paper, we apply the SPLS framework to the following reaction–diffusion discretization problem

$$\begin{cases} -\varepsilon\, \Delta u + cu = f & \text{in } \ \Omega, \\ \qquad\quad u = 0 & \text{on } \partial\Omega, \end{cases} \tag{1.1}$$

for non-negative constants ε and $c(x) \geq c_0 > 0$ on Ω, a bounded domain in $\mathbb{R}^d$. A particular problem of interest is the reaction dominated case in which $\varepsilon \ll 1$. These types of equations arise in heat transfer problems in thin domains [5] as well as when using small step sizes in implicit time discretizations of parabolic reaction–diffusion type problems [6]. The solutions to these problems are characterized by exponential boundary layers of width $\mathcal{O}(\varepsilon^{1/2}\ln(1/\varepsilon))$ [7], which pose numerical challenges due to the ε-dependence of the ellipticity constant.

Finite element methods for these type of problems have been intensively studied, see e.g., [6–9,9–15]. Some of these references include least-squares approaches. In [6], a mixed method approach is given by introducing a new variable for ∇u, rewriting (1.1) as a first order system, and utilizing $H(\mathrm{div};\Omega)$ conforming spaces. We consider an approach in which

[☆] The work was supported by NSF, DMS-1522454.

[*] Corresponding author.
E-mail addresses: bacuta@udel.edu (C. Bacuta), jacavagj@lafayette.edu (J. Jacavage).

https://doi.org/10.1016/j.rinam.2020.100105

we adopt a mixed formulation, the use of graph type trial spaces, and the adoption of an optimal trial norm to obtain stability independent of the parameter ε.

The paper is organized as follows. In Section 2, we introduce the notation and review the SPLS approach. Section 3 details the steps to fit (1.1) into the SPLS framework. Section 4 involves the discretization and choices of discrete trial spaces using a piecewise linear test space. The stability of the proposed discrete spaces is discussed in Section 5. In Section 6, we describe the construction of a Shishkin mesh, which is a specific type of mesh used to resolve the boundary layers exhibited by the solutions for small ε. Lastly, numerical results are given in Section 7 to support and show the performance of the SPLS approach for the reaction–diffusion problem.

2. The notation and the general SPLS approach

We now review the main ideas and concepts for the SPLS nonconforming discretion of a general mixed variational formulation.

2.1. The abstract variational formulation at the continuous level

We consider the following general Petrov–Galerkin formulation: Given $F \in V^*$, find $p \in Q$ such that

$$b(v,p) = \langle F, v\rangle \quad \text{for all } v \in V, \tag{2.1}$$

where V and Q are Hilbert spaces and $b(\cdot,\cdot)$ is a continuous bilinear form on $V \times Q$ satisfying an inf − sup condition. We assume the inner products $a_0(\cdot,\cdot)$ and $(\cdot,\cdot)_{\tilde{Q}}$ induce the norms $|\cdot|_V = |\cdot| = a_0(\cdot,\cdot)^{1/2}$ and $\|\cdot\|_{\tilde{Q}} = \|\cdot\| = (\cdot,\cdot)_{\tilde{Q}}^{1/2}$. We denote the dual of V by V^* and the dual pairing on $V^* \times V$ by $\langle\cdot,\cdot\rangle$. We view Q, the trial space in (2.1), as a subspace of larger (host) space $\tilde{Q}$ and equip Q with the induced inner product and norm from $\tilde{Q}$. The extra space $\tilde{Q}$ is needed for the SPLS *non-conforming discretization*. We assume that $b(\cdot,\cdot)$ is a continuous bilinear form on $V \times \tilde{Q}$ satisfying the following sup − sup condition on $V \times \tilde{Q}$ and inf − sup condition on $V \times Q$,

$$\sup_{p\in\tilde{Q}}\sup_{v\in V}\frac{b(v,p)}{|v|\,\|p\|} = M < \infty, \quad \text{and} \quad \inf_{p\in Q}\sup_{v\in V}\frac{b(v,p)}{|v|\,\|p\|} = m > 0. \tag{2.2}$$

With the form b, we associate the operator $B : V \to \tilde{Q}$ defined by

$$(Bv, q)_{\tilde{Q}} = b(v,q) \quad \text{for all } v \in V, q \in Q.$$

It is well known that if a bounded form $b : V \times \tilde{Q} \to \mathbb{R}$ satisfies (2.2) and the data $F \in V^*$ satisfies the *compatibility condition*

$$\langle F, v\rangle = 0 \quad \text{for all } v \in V_0 := \{v \in V\,|b(v,q) = 0, \quad \text{for all } q \in Q\}, \tag{2.3}$$

then the problem (2.1) has a unique solution, see e.g. [16,17]. It was mentioned in a few papers, [4,18–20], that solving the mixed problem (2.1) reduces to solving a standard saddle point formulation: Find $(w,p) \in (V,Q)$ such that

$$\begin{array}{llll} a_0(w,v) & + \quad b(v,p) & = \langle F, v\rangle & \text{for all } v \in V, \\ b(u,q) & & = 0 & \text{for all } q \in Q. \end{array} \tag{2.4}$$

In fact, we have that p is the unique solution of (2.1) *if and only if* $(w = 0, p)$ solves (2.4), and the result remains valid if the form $a_0(\cdot,\cdot)$ in (2.4) is replaced by any other symmetric bilinear form on V that leads to an equivalent norm on V.

2.2. The concept of optimal test norm

If we assume that $Range(B) \subset Q$ and that the operator $B : V \to Q$ is injective ($V_0 = Ker(B) = \{0\}$) then, as in [20–25], we can define an equivalent norm on V, that is operator dependent, by

$$|v|_{opt} := \sup_{p\in Q}\frac{b(v,p)}{\|p\|} = \sup_{p\in Q}\frac{(Bv,p)}{\|p\|} = \|Bv\|_Q = \|Bv\|_{\tilde{Q}}.$$

We will refer to this as the *optimal test norm*. By replacing the form $a_0(\cdot,\cdot)$ in (2.4) with the inner product induced by the *optimal test norm*, i.e., $a_{opt}(u,v) := (Bu, Bv)_Q$, we obtain that the Schur complement of the new saddle point system becomes the identity, and both the continuity constant M and the inf − sup constant m are equal to 1. Thus, the stability of the new saddle point formulation is optimal.

2.3. The abstract variational formulation at the discrete level

The non-conforming (trial space) *SPLS discretization* of (2.1) is defined as a saddle point discretization of (2.4) with $V_h \subset V$ and with $\mathcal{M}_h$ a subspace of $\tilde{Q}$, but in general *not necessarily a subspace of* Q. Assume that the following discrete sup − sup and inf − sup conditions hold for the pair $(V_h, \mathcal{M}_h)$:

$$\sup_{p_h\in\mathcal{M}_h} \sup_{v_h\in V_h} \frac{b(v_h, p_h)}{|v_h| \, \|p_h\|} = M_h \leq M \text{ and} \tag{2.5}$$

$$\inf_{p_h\in\mathcal{M}_h} \sup_{v_h\in V_h} \frac{b(v_h, p_h)}{|v_h| \, \|p_h\|} = m_h > 0. \tag{2.6}$$

The discrete mixed variational formulation of (2.1) is: Find $p_h \in \mathcal{M}_h$ such that

$$b(v_h, p_h) = \langle F, v_h\rangle \qquad \text{for all } v_h \in V_h. \tag{2.7}$$

In general, this problem might not have unique solution. However, it is well known that the discrete saddle point variational (re)formulation: Find $(w_h, p_h) \in V_h \times \mathcal{M}_h$ such that

$$\begin{array}{llll} a_0(w_h, v_h) & + \quad b(v_h, p_h) & = \langle F, v_h\rangle & \text{for all } v_h \in V_h, \\ b(w_h, q_h) & & = 0 & \text{for all } q_h \in \mathcal{M}_h, \end{array} \tag{2.8}$$

has a unique solution. The variational formulation (2.8) is the *non-conforming saddle point least squares* (n-c SPLS) *discretization* of (2.1). In what follows, $V_h \subset V$ will be chosen as a standard conforming finite element space. On the other hand, each choice of the space $\mathcal{M}_h$, possibly non-conforming to Q, leads to a new SPLS discretization for which $p_h \in \mathcal{M}_h \subset \tilde{Q}$. The discrete operator associated with the form $a_0(\cdot, \cdot)$ on V_h is $A_h : V_h \to V_h^*$, and the discrete linear operators $B_h : V_h \to \mathcal{M}_h$ and $B_h^* : \mathcal{M}_h \to V_h^*$ are defined by

$$(B_h v_h, q_h)_{\mathcal{M}_h} = b(v_h, q_h) = \langle B_h^* q, v_h\rangle \quad \text{for all } v_h \in V_h, \ q \in \mathcal{M}_h.$$

The Schur complement of (2.8) is denoted by $S_h = B_h A_h^{-1} B_h^*$.

Remark 2.1. Note that by using the definition of $B_h w_h$, with $w_h \in V_h$, the second equation in (2.8) is equivalent to

$$B_h \, w_h = 0.$$

2.4. Choosing trial spaces

Let V_h be a *finite element subspace* of V. As presented in [1,3], using the current notation, we provide two types of general trial spaces $\mathcal{M}_h$ that can be considered for the SPLS discretization. The first choice for $\mathcal{M}_h$, the *no projection trial space*, can be viewed as a conforming trial space and has already been investigated in [2,4,19]. We review the *no projection trial space* here because it helps with the analysis of the second choice of $\mathcal{M}_h$, the *projection trial space*.

2.4.1. No projection (conforming) trial space

We first consider the case when $\tilde{Q} = Q$ and $\mathcal{M}_h$ is given by

$$\mathcal{M}_h := BV_h \subset Q = \tilde{Q}.$$

In this case, we that $V_{h,0} \subset V_0$, where $V_{h,0} := Ker(B_h)$. As presented in [1], a discrete inf − sup condition holds, i.e.,

$$m_{h,0} := \inf_{p_h\in\mathcal{M}_h} \sup_{v_h\in V_h} \frac{b(v_h, p_h)}{|v_h| \ \|p_h\|} > 0. \tag{2.9}$$

Thus, we have that (2.8) has a unique solution $(0, p_h) \in (V_h, \mathcal{M}_h)$.

In addition, see [1], if p is the solution of (2.1), then

$$\|p - p_h\| = \inf_{q_h\in\mathcal{M}_h} \|p - q_h\|,$$

i.e., p_h is the orthogonal projection of p onto $\mathcal{M}_h$. Moreover, if the inner product $a_0(\cdot, \cdot)$ in (2.8) is replaced by $a_{opt}(u, v) = (Bu, Bv)_Q$, i.e., we choose the *optimal test norm*, it is easy to check that $M_h = m_{h,0} = 1$. Thus, in this case, we have optimal discrete stability and optimal approximability.

2.4.2. Projection type trial space

Let $\tilde{\mathcal{M}}_h \subset \tilde{Q}$ be a finite dimensional subspace equipped with the inner product $(\cdot, \cdot)_h$. The corresponding induced norm on $\tilde{\mathcal{M}}_h$ will be denoted by $\| \cdot \|_h$. Define the representation operator $R_h : \tilde{Q} \to \tilde{\mathcal{M}}_h$ by

$$(R_h p, q_h)_h := (p, q_h)_{\tilde{Q}} \quad \text{for all } q_h \in \tilde{\mathcal{M}}_h.$$

Here, $R_h p$ is the Riesz representation of $p \to (p, q_h)_{\tilde{Q}}$ as a functional on $(\tilde{\mathcal{M}}_h, (\cdot, \cdot)_h)$. In the case when $(\cdot, \cdot)_h$ coincides with the inner product on $\tilde{Q}$, we have that R_h is precisely the orthogonal projection onto $\tilde{\mathcal{M}}_h$.

Since the space $\tilde{\mathcal{M}}_h$ is finite dimensional, there exist constants k_1, k_2 such that

$$k_1 \|q_h\| \le \|q_h\|_h \le k_2 \|q_h\| \quad \text{for all } q_h \in \tilde{\mathcal{M}}_h. \tag{2.10}$$

We further assume that the equivalence is uniform with respect to h, i.e., the constants k_1, k_2 are independent of h. Using the operator R_h, we define $\mathcal{M}_h$ as

$$\mathcal{M}_h := R_h B V_h \subset \tilde{\mathcal{M}}_h \subset \tilde{Q}.$$

Remark 2.2. We note that by using the definitions of B, B_h, R_h and $\mathcal{M}_h$, for any $v_h \in V_h$ and $q_h \in \mathcal{M}_h$, we have

$$(B_h v_h, q_h)_{\mathcal{M}_h} = b(v_h, q_h) = (B v_h, q_h)_{\tilde{Q}} = (R_h B v_h, q_h)_{\mathcal{M}_h}.$$

Thus,

$$B_h v_h = R_h B v_h, \qquad \text{for all } v_h \in V_h, \tag{2.11}$$

and using Remark 2.1 for this choice of trial space, the second equation in (2.8) is equivalent to

$$R_h B\, w_h = 0.$$

The following proposition gives a sufficient condition on R_h to ensure the discrete inf − sup condition is satisfied and relates the stability of the families of spaces $\{(V_h, BV_h)\}$ and $\{(V_h, R_h B V_h)\}$. The result was proved in [1].

Proposition 2.3. *Assume that*

$$\|R_h q_h\|_h \ge \tilde{c}\, \|q_h\| \qquad \textit{for all } q_h \in BV_h, \tag{2.12}$$

with a constant $\tilde{c}$ independent of h. Then

$$\inf_{p_h \in \mathcal{M}_h} \sup_{v_h \in V_h} \frac{b(v_h, p_h)}{|v_h|\ \|p_h\|_h} \ge \tilde{c}\, m_{h,0} > 0, \tag{2.13}$$

where $m_{h,0}$ is defined in (2.9).

As a consequence of Proposition 2.3, we have that (2.8) has a unique solution $(0, p_h) \in (V_h, \mathcal{M}_h)$.

Regarding the *approximability property* of the projection type trial space, the following proposition was proved in [1].

Proposition 2.4. *If p is the solution of (2.1) and p_h is the second component of the solution of (2.8), then*

$$\|p - p_h\| \le C \inf_{q_h \in \mathcal{M}_h} \|p - q_h\|,$$

with $C = 1 + \frac{1}{k_1 \tilde{c}}$, where k_1 and $\tilde{c}$ were introduced in (2.10) and (2.12), respectively.

Remark 2.5. The no projection trial space described in Section 2.4.1 can be viewed as the special case of the projection type trial space when $R_h = I$ and the inner product $(\cdot, \cdot)_h$ on $\mathcal{M}_h$ is taken to be the original inner product on $\tilde{Q} = Q$. Thus, in what follows, we will consider $\mathcal{M}_h$ to be equipped with the inner product $(\cdot, \cdot)_h$ for both the no projection and projection type trial spaces.

2.5. An Uzawa CG iterative solver

Note that a global linear system may be difficult to assemble when solving (2.8) on $(V_h, \mathcal{M}_h = R_h B V_h)$, especially if the operator R_h involves a nonlocal projection. In this case, bases for the trial spaces $\mathcal{M}_h$ might be difficult to find. One can solve (2.8) and avoid building a basis for $\mathcal{M}_h$ by using an Uzawa type algorithm. To simplify the presentation, we will focus on the Uzawa Conjugate Gradient (UCG) algorithm. Other Uzawa type algorithms are discussed in [1].

Algorithm 2.6 (*(UCG) Algorithm*)**.**

Step 1: Choose any $p_0 \in \mathcal{M}_h$. **Compute** $w_1 \in V_h$, $q_1, d_1 \in \mathcal{M}_h$ by

$$\begin{aligned} a_0(w_1, v_h) &= \langle f_h, v_h \rangle - b(v, p_0) && \text{for all } v_h \in V_h, \\ (q_1, q)_h &= b(w_1, q) && \text{for all } q \in \mathcal{M}_h, \qquad d_1 := q_1. \end{aligned}$$

Step 2: For $j = 1, 2, \dots,$ **compute** $h_j, \alpha_j, p_j, w_{j+1}, q_{j+1}, \beta_j, d_{j+1}$ by

$$\begin{aligned}
&(\textbf{UCG1}) && a_0(h_j, v_h) = -\, b(v_h, d_j) \quad \text{for all } v_h \in V_h\\
&(\textbf{UCG}\alpha) && \alpha_j = -\,\frac{(q_j, q_j)_h}{b(h_j, q_j)}\\
&(\textbf{UCG2}) && p_j = p_{j-1} + \alpha_j\, d_j\\
&(\textbf{UCG3}) && w_{j+1} = w_j + \alpha_j\, h_j\\
&(\textbf{UCG4}) && (q_{j+1}, q)_h = b(w_{j+1}, q) \quad \text{for all } q \in \mathcal{M}_h\\
&(\textbf{UCG}\beta) && \beta_j = \frac{(q_{j+1}, q_{j+1})_h}{(q_j, q_j)_h}\\
&(\textbf{UCG6}) && d_{j+1} = q_{j+1} + \beta_j d_j.
\end{aligned}$$

Remark 2.7. From **(UCG4)**, we have that $q_{j+1} = B_h w_{j+1}$. Using (2.11) of Remark 2.2, we have that $q_{j+1} = R_h B w_{j+1}$. Thus, from the definitions of the operators R_h and B, for any q in the possibly larger space $\tilde{\mathcal{M}_h}$, we have

$$(q_{j+1}, q)_h = (R_h B w_{j+1}, q)_h = (B w_{j+1}, q)_{\tilde{Q}} = b(w_j, q).$$

This implies that q_{j+1} can be computed by inverting the Gram matrix corresponding to a basis of $\tilde{\mathcal{M}_h}$ (which in our applications is component-wise a space of continuous piecewise linear functions), and the Gram matrix corresponding to a basis of $\mathcal{M}_h = R_h B V_h$ is not needed for the computation of q_{j+1} in **(UCG4)** or of q_1 in **Step 1**.

The main inversion needed at each step involves $a_0(\cdot, \cdot)$ in

Step 1 or (**UCG1**). In operator form, these steps become

$$w_1 = A_h^{-1}(f_h - B_h^* p_0), \qquad \text{and} \qquad h_j = -A_h^{-1}(B_h^* d_j). \tag{2.14}$$

Regarding the convergence of the UCG algorithm, it is well known that if (w_h, p_h) is the discrete solution of (2.8) and (w_{j+1}, p_j) is the jth iteration for the UCG algorithm, then $(w_{j+1}, p_j) \to (w_h, p_h)$. The rate of convergence depends on the condition number of the Schur complement S_h, and $\|q_j\|_h = (q_j, q_j)_h^{1/2}$ is an optimal estimator for the iteration error $\|p_h - p_j\|_h$, see [1]. In implementation, if an a-priori estimate for the discretization error $\|p - p_h\|_{\tilde{Q}} \approx h^\alpha$ is available, we can use it to match the iteration error by imposing the following stopping criterion for the UCG:

$$\|q_j\|_h \le c_0 h^\alpha. \tag{2.15}$$

3. SPLS for reaction–diffusion equations

In this section, we will describe how to apply the general SPLS theory to problem (1.1). A standard variational formulation for (1.1) is: Find $u \in H_0^1(\Omega)$ such that

$$(\varepsilon \nabla u, \nabla v) + (cu, v) = (f, v) \quad \text{for all } v \in H_0^1(\Omega). \tag{3.1}$$

In what follows, $(\cdot, \cdot)$ and $\|\cdot\|$ will denote the standard L^2 inner product and norm, respectively. To fit this equation into the SPLS framework, we let $V := H_0^1(\Omega)$, $\tilde{Q} := L^2(\Omega) \times L^2(\Omega)^d$, and Q be the graph of the operator $\varepsilon\nabla : H_0^1(\Omega) \to L^2(\Omega)^d$, i.e.,

$$Q := G(\varepsilon\nabla) = \left\{ \left(\begin{smallmatrix} v \\ \varepsilon\nabla v \end{smallmatrix}\right) \mid v \in H_0^1(\Omega) \right\}.$$

Since the operator $\varepsilon\nabla$ is bounded from $H_0^1(\Omega)$ to $L^2(\Omega)^d$, the space Q is closed by the Closed Graph Theorem. We define the bilinear form $b : V \times \tilde{Q} \to \mathbb{R}$ as

$$b\left(v, \left(\begin{smallmatrix} q \\ \mathbf{q} \end{smallmatrix}\right)\right) := (cq, v) + (\mathbf{q}, \nabla v) \quad \text{for all } v \in V, \left(\begin{smallmatrix} q \\ \mathbf{q} \end{smallmatrix}\right) \in \tilde{Q},$$

and the linear functional $F \in V^*$ as

$$\langle F, v \rangle := (f, v) \quad \text{for all } v \in H_0^1(\Omega).$$

With this setting, the SPLS formulation of (3.1) is: Find $\mathbf{p} = \left(\begin{smallmatrix} u \\ \varepsilon\nabla u \end{smallmatrix}\right) \in Q$ such that

$$b(v, \mathbf{p}) = (cu, v) + (\varepsilon\nabla u, \nabla v) = (f, v) \quad \text{for all } v \in V. \tag{3.2}$$

On V, we consider first the standard the inner product defined by

$$a_0(u, v) = (\nabla u, \nabla v) \quad \text{for all } u, v \in V,$$

On $\tilde{Q}$, we consider the weighted inner product

$$\left(\left(\begin{smallmatrix} q \\ \mathbf{q} \end{smallmatrix}\right), \left(\begin{smallmatrix} p \\ \mathbf{p} \end{smallmatrix}\right)\right)_{\tilde{Q}} = (cq, p) + (\varepsilon^{-1}\mathbf{q}, \mathbf{p}) := (q, p)_c + (\mathbf{q}, \mathbf{p})_{\varepsilon^{-1}}, \left(\begin{smallmatrix} q \\ \mathbf{q} \end{smallmatrix}\right), \left(\begin{smallmatrix} p \\ \mathbf{p} \end{smallmatrix}\right) \in \tilde{Q}. \tag{3.3}$$

The corresponding norm is

$$\| \left(\begin{smallmatrix} q \\ \mathbf{q} \end{smallmatrix} \right) \|_{\tilde{Q}} = \left(\|c^{1/2} q\|^2 + \|\varepsilon^{-1/2} \mathbf{q}\|^2 \right)^{1/2}.$$

The operator $B : V \to Q$ is given by

$$Bv = \left(\begin{smallmatrix} v \\ \varepsilon \nabla v \end{smallmatrix} \right) \quad \text{for all } v \in V.$$

Thus, the *optimal test norm* on V is induced by the inner product

$$a_{opt}(u, v) = (Bu, Bv)_Q = (\varepsilon \nabla u, \nabla v) + (cu, v) \quad \text{for all } u, v \in V,$$

which gives rise to the norm

$$|v|_{opt} = \left(\|c^{1/2} v\|^2 + \|\varepsilon^{1/2} \nabla v\|^2 \right)^{1/2}.$$

The compatibility condition (2.3) is automatically satisfied as

$$V_0 = \mathrm{Ker}(B) = \{v \in H_0^1(\Omega) \mid Bv = 0\} = \{0\}.$$

In addition, according to Section 2.2 we have $M = m = 1$. One can also directly check that

$$\sup_{v \in V} \frac{b(v, \left(\begin{smallmatrix} u \\ \varepsilon \nabla u \end{smallmatrix} \right))}{|v|_{opt}} = \| \left(\begin{smallmatrix} u \\ \varepsilon \nabla u \end{smallmatrix} \right) \|_Q, \tag{3.4}$$

for any $\left(\begin{smallmatrix} u \\ \varepsilon \nabla u \end{smallmatrix} \right) \in Q$. This leads to optimal continuity and inf − sup constants. Thus, the variational problem (3.2) is suitable for SPLS discretization. In the implementation of UCG, to take advantage of the uniform stability, we replace $a_0(\cdot, \cdot)$ by $a_{opt}(\cdot, \cdot)$.

4. SPLS discretization for reaction–diffusion problems

In this section, we will discuss possible choices for the discrete spaces as well as their stability. The choices for the trial space will be based on the *no projection* and *projection* type spaces outlined in Sections 2.4.1 and 2.4.2. For the discrete test space, we take $V_h \subset V = H_0^1(\Omega)$ to be the space of continuous piecewise polynomials of degree k with respect to the mesh $\mathcal{T}_h$.

4.1. No projection trial space

Following Section 2.4.1, we consider the case when the trial space $\mathcal{M}_h$ is given by

$$\mathcal{M}_h := BV_h = \begin{pmatrix} I \\ \varepsilon \nabla \end{pmatrix} V_h,$$

where $I : V_h \to V_h$ is the identity operator and the inner product is chosen to coincide with the inner product on $\tilde{Q}$. By a similar argument used to show (3.4), or according to Section 2.2, we obtain

$$\sup_{v_h \in V_h} \frac{b\left(v_h, \left(\begin{smallmatrix} u_h \\ \varepsilon \nabla u_h \end{smallmatrix} \right)\right)}{|v_h|_V} = \left\| \left(\begin{smallmatrix} u_h \\ \varepsilon \nabla u_h \end{smallmatrix} \right) \right\|_{\tilde{Q}}, \tag{4.1}$$

for any $\left(\begin{smallmatrix} u_h \\ \varepsilon \nabla u_h \end{smallmatrix} \right) \in \mathcal{M}_h$. Thus, we do have stability in this case. Furthermore, the stability constant $m_{h,0}$ and the boundness constant M_h are independent of both the parameters h and ε.

The discrete mixed variational formulation in this case becomes: Find $\mathbf{p}_h = \left(\begin{smallmatrix} u_h \\ \varepsilon \nabla u_h \end{smallmatrix} \right)$, with $u_h \in V_h$, such that

$$b(v_h, \mathbf{p}_h) = (\varepsilon \nabla u_h, \nabla v_h) + (cu_h, v_h) = (f, v_h) \quad \text{for all } v_h \in V_h.$$

The discrete saddle point reformulation to be solved is: Find $\left(w_h, \mathbf{p}_h = \left(\begin{smallmatrix} u_h \\ \varepsilon \nabla u_h \end{smallmatrix} \right)\right)$ such that

$$\begin{array}{lcll} \varepsilon(\nabla w_h + \nabla u_h, \nabla v_h) & + & c(w_h + u_h, v_h) & = (f, v_h) \qquad \text{for all } v_h \in V_h, \\ \left(\begin{smallmatrix} w_h \\ \varepsilon \nabla w_h \end{smallmatrix} \right) & & & = \mathbf{0}. \end{array} \tag{4.2}$$

Remark 4.1. In this case, the UCG algorithm for (4.2) converges in one iteration. However, this case is not of practical interest as a standard Galerkin method for the original problem on the same V_h would produce the same result. We consider this case for the theoretical purposes of addressing the stability and discretization when using the *projection trial space*, which is presented in the next subsection. In addition, this case acts as a preliminary test for the *projection trial space* discretization.

4.2. Projection type trial space

For the projection type trial space, we first define $\tilde{\mathcal{M}}_h \subset \tilde{Q} = L^2(\Omega) \times L^2(\Omega)^d$ to be

$$\tilde{\mathcal{M}}_h := M_{h,0} \times \varepsilon \mathbf{M}_{h,0},$$

where $M_{h,0}$ consists of continuous piecewise polynomials of degree k with respect to the mesh $\mathcal{T}_h$ with no restrictions on the boundary. The space $\mathbf{M}_{h,0}$ is the vector-valued product space in which each component consists of continuous piecewise polynomials of degree k. Two different choices for the projection type trial space, based on the inner product chosen for $\tilde{\mathcal{M}}_h$, are given in the previous section. The first is outlined in this section. The second is outlined in Section 5.1.

For the first type of projection trial space, we equip $\tilde{\mathcal{M}}_h$ with the inner product induced from $\tilde{Q}$ and define $R_h \begin{pmatrix} q \\ \mathbf{q} \end{pmatrix}$ as the orthogonal projection of $\begin{pmatrix} q \\ \mathbf{q} \end{pmatrix}$ onto $\tilde{\mathcal{M}}_h$ with respect to the $(\cdot,\cdot)_{\tilde{Q}}$ inner product. More specifically, we have that

$$R_h \begin{pmatrix} q \\ \mathbf{q} \end{pmatrix} = \begin{pmatrix} Q_h^1 q \\ Q_h^2 \mathbf{q} \end{pmatrix},$$

where $Q_h^1 : L^2(\Omega) \to M_{h,0}$ is the orthogonal projection with respect to the weighted inner product $(\cdot,\cdot)_c$ and $Q_h^2 : L^2(\Omega)^d \to \mathbf{M}_{h,0}$ is the orthogonal projection with respect to the weighted inner product $(\cdot,\cdot)_{\varepsilon^{-1}}$, where $(\cdot,\cdot)_c$ and $(\cdot,\cdot)_{\varepsilon^{-1}}$ are defined in (3.3). We now define the projection type trial space as

$$\mathcal{M}_h := R_h^{\text{orth}} B V_h,$$

where the elements are given by

$$R_h^{\text{orth}} B v_h = \begin{pmatrix} Q_h^1 v_h \\ Q_h^2(\varepsilon \nabla v_h) \end{pmatrix}.$$

Remark 4.2. In general, $\mathcal{M}_h$ constructed in this way is not contained in Q due to the fact that the range of the projection Q_h^2 might not be a gradient field. This justifies here the choice of a larger host space $\tilde{Q}$ for the discrete trial space $\mathcal{M}_h$.

The discrete mixed variational formulation in this case is: Find $\mathbf{p}_h = R_h^{\text{orth}} B u_h$, with $u_h \in V_h$, such that

$$b(v_h, \mathbf{p}_h) = (f, v_h) \quad \text{for all } v_h \in V_h,$$

where $b(\cdot,\cdot)$ is defined in Section 3. The SPLS discretization of (3.1), with *optimal test norm* is: Find $(w_h, \mathbf{p}_h = R_h^{\text{orth}} A \nabla u_h) \in V_h \times \mathcal{M}_h$ such that

$$\begin{array}{llll} a_{opt}(w_h, v_h) & + \quad b(v_h, \mathbf{p}_h) & = (f, v_h) & \quad \text{for all } v_h \in V_h, \\ R_h^{\text{orth}} B w_h & & = \mathbf{0}. & \end{array} \tag{4.3}$$

5. Piecewise linear test space

In this section, we discuss the stability for the family of spaces $\{(V_h, \mathcal{M}_h)\}$, where $\mathcal{M}_h$ is as outlined in Section 4.2. For simplicity, we assume $\Omega \subset \mathbb{R}^2$ is a polygonal domain. The results can be extended to polyhedral domains in $\mathbb{R}^3$. We also assume that the triangular mesh $\mathcal{T}_h$ is locally quasi-uniform. Let $\{z_1, \dots, z_N\}$ be the set of all nodes of $\mathcal{T}_h$ and assume all triangles adjacent to z_j are of regular shape and their area is of order h_j^2. In this notation, the mesh size of $\mathcal{T}_h$ is $h := \max\{h_1, h_2, \dots, h_N\}$.

We take V_h to be the space consisting of piecewise linear polynomials with respect to $\mathcal{T}_h$ vanishing on the boundary of Ω. Also, we take $M_{h,0}$ to consist of continuous linear piecewise polynomials with respect to the mesh $\mathcal{T}_h$. Let $\{\phi_1, \dots, \phi_N\}$ denote a nodal basis for $M_{h,0}$ with respect to the mesh $\mathcal{T}_h$ and $\{\mathbf{\Phi}_1, \dots, \mathbf{\Phi}_{2N}\}$ denote a nodal basis for $\mathbf{M}_{h,0}$, where $\mathbf{\Phi}_j = (\phi_j, 0)^T$ and $\mathbf{\Phi}_{N+j} = (0, \phi_j)^T$ for $j = 1, \dots, N$. With this notation, $\{\phi_j\}_{j=1}^N \cup \{\varepsilon \mathbf{\Phi}_j\}_{j=1}^{2N}$ is a basis for $\tilde{\mathcal{M}}_h$. We further define M_ε to be the matrix whose entries are $(\varepsilon \mathbf{\Phi}_i, \varepsilon \mathbf{\Phi}_j)_{\varepsilon^{-1}} = (\varepsilon \mathbf{\Phi}_i, \mathbf{\Phi}_j)$ and $H := \operatorname{diag}\left(h_1^2, h_2^2, \dots, h_N^2\right)$. Lastly, we let

$$D_\varepsilon = \left[\begin{array}{c|c} \varepsilon H & \\ \hline & \varepsilon H \end{array}\right].$$

In what follows, $\langle \cdot,\cdot \rangle_e$ denotes the standard euclidean inner product.

Lemma 5.1. *Under the assumptions of Section 5, there exists a constant C independent of h and ε such that*

$$\langle M_\varepsilon \boldsymbol{\gamma}, \boldsymbol{\gamma} \rangle_e \le C \, \langle D_\varepsilon \boldsymbol{\gamma}, \boldsymbol{\gamma} \rangle_e \quad \textit{for all } \boldsymbol{\gamma} \in \mathbb{R}^{2N}. \tag{5.1}$$

Consequently,

$$\langle M_\varepsilon^{-1} \boldsymbol{\gamma}, \boldsymbol{\gamma} \rangle_e \ge C \langle D_\varepsilon^{-1} \boldsymbol{\gamma}, \boldsymbol{\gamma} \rangle_e \quad \textit{for all } \boldsymbol{\gamma} \in \mathbb{R}^{2N}. \tag{5.2}$$

Proof. Let $\boldsymbol{\gamma} \in \mathbb{R}^{2N}$ and define $\mathbf{q}_h := \sum_{j=1}^{2N} \boldsymbol{\gamma}_j \boldsymbol{\Phi}_j$. Note that

$$\langle M_\varepsilon \boldsymbol{\gamma}, \boldsymbol{\gamma}\rangle_e = (\varepsilon \mathbf{q}_h, \mathbf{q}_h) = \|\varepsilon \mathbf{q}_h\|^2_{\varepsilon^{-1}} = \sum_{\tau \in \mathcal{T}_h} \|\varepsilon \mathbf{q}_h\|^2_{\tau,\varepsilon^{-1}}. \tag{5.3}$$

If $\tau = [z_{1_\tau}, z_{2_\tau}, z_{3_\tau}]$, then

$$\mathbf{q}_h\Big|_\tau = \begin{pmatrix} \sum_{j=1}^{3} \gamma_{j_\tau} \phi_{j_\tau} \\ \sum_{j=1}^{3} \gamma_{(j+N)_\tau} \phi_{j_\tau} \end{pmatrix}.$$

Hence,

$$\|\varepsilon \mathbf{q}_h\|^2_{\tau,\varepsilon^{-1}} \le C\,|\tau| \left(\varepsilon \sum_{j=1}^{3} \gamma^2_{j_\tau} + \varepsilon \sum_{j=1}^{3} \gamma^2_{(j+N)_\tau} \right). \tag{5.4}$$

Using (5.3), (5.4), and the fact that each coefficient γ_k can repeat at most three times, we obtain

$$\langle M_\varepsilon \boldsymbol{\gamma}, \boldsymbol{\gamma}\rangle_e \le C \left(\varepsilon \sum_{j=1}^{N} h_j^2 \gamma_j^2 + \varepsilon \sum_{j=1}^{N} h_j^2 \gamma_{j+N}^2 \right) = C\ \langle D_\varepsilon \boldsymbol{\gamma}, \boldsymbol{\gamma}\rangle_e\,.$$

Estimate (5.2) follows from (5.1). □

We now show that (2.12) is satisfied for the operator R_h^{orth} defined Section 4.2.

Lemma 5.2. *Under the assumptions of Section 5, there exists a constant $\tilde{C}$, independent of h and ε, such that*

$$\|R_h^{\text{orth}} \left(\begin{smallmatrix} v_h \\ \varepsilon\nabla v_h \end{smallmatrix} \right) \|_h \ge \tilde{C}\, \| \left(\begin{smallmatrix} v_h \\ \varepsilon\nabla v_h \end{smallmatrix} \right) \|_{\tilde{Q}} \quad \textit{for all } v_h \in V_h. \tag{5.5}$$

Proof. For a fixed $\left(\begin{smallmatrix} v_h \\ \varepsilon\nabla v_h \end{smallmatrix} \right)$, with $v_h \in V_h$, we define the vector $\mathbf{G}_h \in \mathbb{R}^{2N}$ by

$$(G_h)_i := (\varepsilon \nabla v_h, \varepsilon \boldsymbol{\Phi}_i)_{\varepsilon^{-1}} = (\varepsilon \nabla v_h, \boldsymbol{\Phi}_i)\ \ i = 1, \dots, 2N.$$

Recall that

$$R_h^{\text{orth}} \begin{pmatrix} v_h \\ \varepsilon \nabla v_h \end{pmatrix} = \begin{pmatrix} Q_h^1 v_h \\ Q_h^2(\varepsilon \nabla v_h) \end{pmatrix},$$

where Q_h^1 and Q_h^2 are defined in Section 4.2. Note that $Q_h^1 v_h = v_h$ and let

$$Q_h^2(\varepsilon \nabla v_h) = \sum_{i=1}^{2N} \alpha_i \varepsilon \boldsymbol{\Phi}_i\,.$$

Thus, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_{2N})^T$ is a solution to

$$M_\varepsilon\, \boldsymbol{\alpha} = \mathbf{G}_h.$$

Using (5.2), we obtain

$$\begin{aligned} \|R_h^{\text{orth}} \left(\begin{smallmatrix} v_h \\ \varepsilon\nabla v_h \end{smallmatrix} \right) \|_h^2 &= \|c^{1/2} v_h\|^2 + \sum_{i,j=1}^{2N} \alpha_i\, \alpha_j \left(\varepsilon \boldsymbol{\Phi}_i, \boldsymbol{\Phi}_j \right) \\ &= \|c^{1/2} v_h\| + \left\langle M_\varepsilon^{-1} \mathbf{G}_h, \mathbf{G}_h \right\rangle_e \\ &\ge C \left(\|c^{1/2} v_h\| + \left\langle D_\varepsilon^{-1} \mathbf{G}_h, \mathbf{G}_h \right\rangle_e \right). \end{aligned}$$

From the definition of the matrix H, we recall $h_i = h_{i+N}$ for $i = 1, \dots, N$. Thus,

$$\begin{aligned} \left\langle D_\varepsilon^{-1} \mathbf{G}_h, \mathbf{G}_h \right\rangle_e &= \sum_{i=1}^{N} h_i^{-2} \left[\varepsilon \left(\frac{\partial v_h}{\partial x}, \phi_i \right)^2 + \varepsilon \left(\frac{\partial v_h}{\partial y}, \phi_i \right)^2 \right] \\ &= \sum_{i=1}^{N} \sum_{\tau \subset supp(\phi_i)} h_i^{-2} (1, \phi_i)^2_\tau \left[\varepsilon \left| \frac{\partial v_h}{\partial x} \right|^2_\tau + \varepsilon \left| \frac{\partial v_h}{\partial y} \right|^2_\tau \right] \\ &\ge \tilde{C} \|\varepsilon \nabla v_h\|^2_{\varepsilon^{-1}}. \end{aligned}$$

Hence,

$$\|R_h^{\mathrm{orth}}\left(\begin{smallmatrix} v_h \\ \varepsilon\nabla v_h\end{smallmatrix}\right)\|_h^2 \ge \tilde{C}\left(\|c^{1/2}v_h\|^2 + \|\varepsilon\nabla v_h\|_{\varepsilon^{-1}}^2\right) = \tilde{C}\|\left(\begin{smallmatrix} v_h \\ \varepsilon\nabla v_h\end{smallmatrix}\right)\|_{\tilde{Q}}. \quad \square$$

As a consequence of Lemma 5.2, Eq. (4.1) (or the fact that $m_{h,0} = 1$), and Proposition 2.3, we obtain the following result.

Theorem 5.3. *Let $\Omega \subset \mathbb{R}^2$ be a polygonal domain and $\{T_h\}$ be a family of locally quasi-uniform meshes for Ω. For each h, let V_h be the space of continuous linear functions with respect to the mesh $\{\mathcal{T}_h\}$ that vanish on $\partial\Omega$ and $\mathcal{M}_h = R_h^{\mathrm{orth}}BV_h$. Then the family of spaces $\{(V_h, \mathcal{M}_h)\}$ is stable.*

Remark 5.4. We note that while the analysis done in this section assumes that the mesh $\mathcal{T}_h$ is locally quasi-uniform, the Shishkin type mesh, that will be outlined in Section 6, does not satisfy this property. Nevertheless, the analysis can be extended to the case of Shishkin type refinement if we follow closely the ε dependence of the coercivity constant $\tilde{C}$.

5.1. Second type of projection trial space

In this section, we consider an inner product on $\tilde{\mathcal{M}}_h$ that is related with lumping the mass matrix. Let $\left(\begin{smallmatrix} q_h \\ \mathbf{q}_h\end{smallmatrix}\right), \left(\begin{smallmatrix} p_h \\ \mathbf{p}_h\end{smallmatrix}\right) \in \tilde{\mathcal{M}}_h$ be two arbitrary elements. We can write

$$\mathbf{q}_h = \sum_{i=1}^{2N} \alpha_i \varepsilon \mathbf{\Phi}_i, \text{ and } \mathbf{p}_h = \sum_{i=1}^{2N} \beta_i \varepsilon \mathbf{\Phi}_i,$$

for some $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_{2N})$ and $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_{2N})$. We consider the inner product

$$\left(\left(\begin{smallmatrix} q_h \\ \mathbf{q}_h\end{smallmatrix}\right), \left(\begin{smallmatrix} p_h \\ \mathbf{p}_h\end{smallmatrix}\right)\right)_h := (cq_h, p_h) + \sum_{i=1}^{2N} \alpha_i\beta_i(1, \varepsilon\mathbf{\Phi}_i),$$

on $\tilde{\mathcal{M}}_h$, where $(\cdot,\cdot)$ represents the standard L^2 inner product. For simplicity, we will denote

$$(\mathbf{q}_h, \mathbf{p}_h)_{\mathrm{lump}} := \sum_{i=1}^{2N} \alpha_i\beta_i(1, \varepsilon\mathbf{\Phi}_i),$$

for the second part of the $(\cdot,\cdot)_h$ inner product. We define $R_h : \tilde{Q} \to \tilde{\mathcal{M}}_h$ by

$$R_h\begin{pmatrix} q \\ \mathbf{q}\end{pmatrix} = \begin{pmatrix} Q_h^1 q \\ Q_h^{\mathrm{lump}}\mathbf{q}\end{pmatrix},$$

where

$$Q_h^{\mathrm{lump}}\mathbf{q} = \sum_{i=1}^{2N} \frac{(\mathbf{q}, \varepsilon\mathbf{\Phi}_i)_{\varepsilon^{-1}}}{(1, \varepsilon\mathbf{\Phi}_i)}\varepsilon\mathbf{\Phi}_i = \sum_{i=1}^{2N} \frac{(\mathbf{q}, \mathbf{\Phi}_i)}{(1, \mathbf{\Phi}_i)}\mathbf{\Phi}_i.$$

We define the projection type trial space in this case as

$$\mathcal{M}_h := R_h^{\mathrm{lump}}BV_h.$$

The problem to be solved using this projection type trial space is identical to (4.3). The following lemma is analogous to 5.2.

Lemma 5.5. *Under the assumptions of Section 5, there exists a constant C, independent of h and ε, such that*

$$\|R_h^{\mathrm{lump}}\left(\begin{smallmatrix} v_h \\ \varepsilon\nabla v_h\end{smallmatrix}\right)\|_h \ge \tilde{C}\,\|\left(\begin{smallmatrix} v_h \\ \varepsilon\nabla v_h\end{smallmatrix}\right)\|_{\tilde{Q}} \quad \textit{for all } v_h \in V_h. \tag{5.6}$$

Proof. Using the same notation from the proof of Lemma 5.2, we obtain

$$\begin{aligned}\|R_h^{\mathrm{lump}}\left(\begin{smallmatrix} v_h \\ \varepsilon\nabla v_h\end{smallmatrix}\right)\|_h^2 &= \|c^{1/2}v_h\|^2 + \sum_{i=1}^{2N} \frac{(\varepsilon\nabla v_h, \varepsilon\mathbf{\Phi}_i)_{\varepsilon^{-1}}^2}{(1, \varepsilon\mathbf{\Phi}_i)^2}(1, \varepsilon\mathbf{\Phi}_i) \\ &= \|c^{1/2}v_h\|^2 + \sum_{i=1}^{2N} \frac{(\varepsilon\nabla v_h, \mathbf{\Phi}_i)^2}{(1, \varepsilon\mathbf{\Phi}_i)} \\ &\ge \tilde{C}\left(\|c^{1/2}v_h\|^2 + \left\langle D_\varepsilon^{-1}\mathbf{G}_h, \mathbf{G}_h\right\rangle_e\right),\end{aligned}$$

where

$$(G_h)_i := (\varepsilon\nabla v_h, \varepsilon\mathbf{\Phi}_i)_{\varepsilon^{-1}} = (\varepsilon\nabla v_h, \mathbf{\Phi}_i) \ \ i = 1, \dots, 2N.$$

From the same techniques used to estimate $\left\langle D_\varepsilon^{-1}\mathbf{G}_h, \mathbf{G}_h\right\rangle_e$ as in the proof of Lemma 5.2, the result follows. □

As a consequence of Lemma 5.5, we obtain the following result.

Theorem 5.6. *Let $\Omega \subset \mathbb{R}^2$ be a polygonal domain and $\{T_h\}$ be a family of locally quasi-uniform meshes for Ω. For each h, let V_h be the space of continuous linear functions with respect to the mesh $\{\mathcal{T}_h\}$ that vanish on $\partial\Omega$ and $\mathcal{M}_h = R_h^{\text{lump}}BV_h$. Then the family of spaces $\{(V_h, \mathcal{M}_h)\}$ is stable.*

6. The construction of a Shishkin mesh

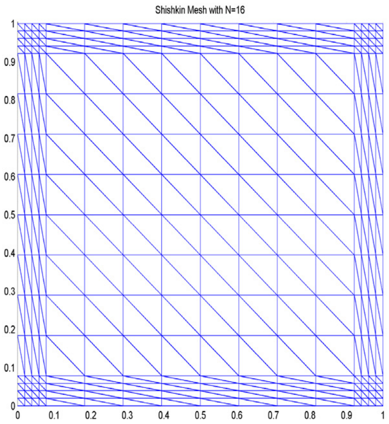
In this section, we describe the construction of a Shishkin mesh [26] for the unit square. These types of meshes are widely used when dealing with reaction dominated diffusion problems in order to resolve the boundary layers exhibited by the solution of the problem. This type of mesh will be used in Sections 7.2, 7.3, and 7.4. We will follow the outline given in [7] to construct a Shishkin mesh for a solution that exhibits boundary layers on all sides of the unit square.

We first assume N is an integer multiple of 8. This parameter will refer to the number of mesh intervals in the x and y directions. The mesh itself is the tensor product of two one-dimensional Shishkin meshes $\mathcal{T}_x \times \mathcal{T}_y$. The process for obtaining $\mathcal{T}_x$ (and $\mathcal{T}_y$) is as follows. The interval $[0, 1]$ is first decomposed into three subintervals $[0, \lambda]$, $[\lambda, 1-\lambda]$, and $[1-\lambda, 1]$, where

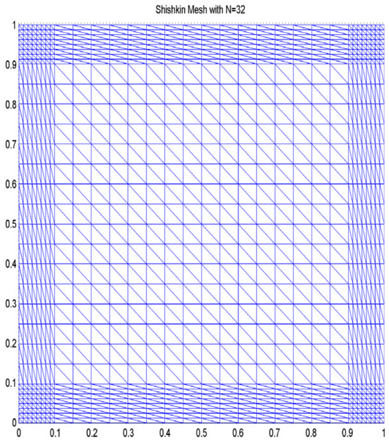
$$\lambda = \min\left\{\frac{1}{4}, 2\sqrt{\frac{\varepsilon}{c^*}}\ln N\right\} \quad \text{with} \ \ 0 < c^* < c. \tag{6.1}$$

The intervals $[0, \lambda]$ and $[1-\lambda, 1]$ are then partitioned into $N/4$ subintervals of length $\frac{4\lambda}{N}$, while the interval $[\lambda, 1-\lambda]$ is partitioned into $N/2$ subintervals of length $\frac{2(1-2\lambda)}{N}$. The triangular mesh is obtained by drawing diagonals from the top left to bottom right of each quadrilateral. The figure below shows an example of the Shishkin mesh generated using $\varepsilon = 10^{-4}$ and $c^* = \sqrt{1/2}$ for $N = 16, 32$, respectively.

Shishkin mesh, $N = 16$



Shishkin mesh, $N = 32$



7. Numerical results

In this section, we present results from applying the SPLS discretization techniques on second order elliptic PDE of the form (1.1). For all of the examples presented, Ω is a bounded polygonal domain, and the test space $V_h \subset H_0^1(\Omega)$ is taken to be the space of continuous piecewise linear polynomials with respect to the Shishkin mesh $\mathcal{T}_h$, unless otherwise noted. We consider all types of trial spaces presented Section 4: the no projection type presented in Section 4.1 and the projection types presented in Sections 4.2 and 5.1. Also, we note that while the theory in this section considers c a non-negative constant, the theory extends to the case where c is a smooth positive function satisfying

$$0 < c_0 \le c(\mathbf{x}) \le c_1 \quad \text{for all } \mathbf{x} \in \Omega,$$

for constants c_0 and c_1.

Table 1
Results for basic unit square example.

Level k	$\mathcal{M}_h = BV_h$			$\mathcal{M}_h = R_h^{\text{orth}} BV_h$			$\mathcal{M}_h = R_h^{\text{lump}} BV_h$		
	Error	Rate	It	Error	Rate	It	Error	Rate	It
1	0.045		1	0.0100		3	0.0202		3
2	0.024	0.903	1	0.0034	1.569	7	0.0090	1.168	6
3	0.012	0.974	1	0.0010	1.735	8	0.0035	1.364	8
4	0.006	0.993	1	3.1e−04	1.724	10	0.0013	1.440	13
5	0.003	0.998	1	8.9e−05	1.785	12	0.0004	1.471	16

For the singularly perturbed problems, we measure the SPLS solution in a balanced norm instead of the norm on $\tilde{Q}$. This is due to the fact that for small ε the L^2 part of the norm (on $\tilde{Q}$) dominates, leading to an unbalanced norm not adequate to accurately measure the error, see [6,7]. More specifically, we measure

$$\text{error} = \left(\|u - u_h\|^2 + \varepsilon^{1/2}\|\nabla u - \nabla u_h\|^2\right)^{1/2},$$

for the no projection type trial space and measure

$$\text{error} = \left(\|u - u_h\|^2 + \varepsilon^{1/2}\|\nabla u - R_h \nabla u_h\|^2\right)^{1/2},$$

for the projection type trial spaces. In the above equation, R_h can be taken as either the orthogonal projection described in Section 4.2 or the lump projection described in Section 5.1.

When using a Shishkin mesh, we used a stopping criterion of

$$\|\mathbf{q}_j\|_h \leq c_0(N^{-1}\ln N),$$

for the no projection type of trial space. This is because standard Galerkin methods for (3.1) obtain a convergence rate of $\mathcal{O}(N^{-1}\ln N)$ using piecewise linear approximation [6,7]. The convergence rates in general when using a Shishkin mesh are computed under the assumption that we have a convergence rate of $\mathcal{O}((N^{-1}\ln N)^r)$. When using a projection type trial space, we used the stopping criterion

$$\|\mathbf{q}_j\|_h \leq c_0(N^{-1}\ln N)^2.$$

7.1. Basic unit square problem

For the first example, we solved (1.1) on the unit square with $c = 1$, $\varepsilon = 1$, and f computed such that the exact solution is given by

$$u(x, y) = x(1 - x)y(1 - y).$$

The family of locally quasi-uniform meshes $\{\mathcal{T}_h\}$ was obtained through a standard uniform refinement strategy starting with a uniform coarse mesh. Here, the mesh size is $h = 2^{-k}$ where k is the level of refinement. Based on the general criterion (2.15), we used a stopping criterion of

$$\|\mathbf{q}_j\|_h \leq c_0 h^2,$$

on each level, and the error is computed in the $\tilde{Q}$ norm. Results for all three types of trial spaces are shown in Table 1. We see $\mathcal{O}(h)$ convergence for the no projection trial space and super-linear convergence for both types of projection type trial spaces.

7.2. Example with boundary layers on all sides

For this example, we solved (1.1) on the unit square with variable coefficient $c = 2(1 + x^2 + y^2)$ and f computed such that the exact solution is given by

$$\begin{aligned} u(x, y) = {} & x(1 - x)\left(1 - e^{-y/\sqrt{\varepsilon}}\right)\left(1 - e^{(y-1)/\sqrt{\varepsilon}}\right) \\ & + y(1 - y)\left(1 - e^{-x/\sqrt{\varepsilon}}\right)\left(1 - e^{(x-1)/\sqrt{\varepsilon}}\right), \end{aligned}$$

as considered in [11]. For this example, the family of Shishkin meshes $\{\mathcal{T}_h\}$ was obtained as in Section 6 with λ in (6.1) computed with $c^* = \sqrt{1/2}$ and the number of subintervals in the x and y directions taken to be $N = 16, 32, 64, 128$, and 256. Table 2 shows results for no projection trial space for a variety of values for ε. We observe $\mathcal{O}(N^{-1}\ln N)$ convergence. Tables 3 and 4 display results for the orthogonal and lump projection type trial spaces. In this case, we observe $\mathcal{O}((N^{-1}\ln N)^2)$ convergence. Furthermore, for all three types of trial spaces we observe the order of convergence is robust with respect to ε.

Table 2
Results for example with boundary layers on all sides and no projection trial space.

$\mathcal{M}_h = BV_h$									
N	$\varepsilon = 1$			$\varepsilon = 10^{-2}$			$\varepsilon = 10^{-4}$		
	Error	Rate	It	Error	Rate	It	Error	Rate	It
16	0.019		1	0.068		1	0.132		1
32	0.009	1.472	1	0.034	1.471	1	0.088	0.854	1
64	0.005	1.356	1	0.017	1.356	1	0.054	0.946	1
128	0.002	1.286	1	0.009	1.285	1	0.032	0.984	1
256	0.001	1.239	1	0.004	1.239	1	0.018	0.996	1
N	$\varepsilon = 10^{-8}$			$\varepsilon = 10^{-12}$			$\varepsilon = 10^{-16}$		
	Error	Rate	It	Error	Rate	It	Error	Rate	It
16	0.133		1	0.134		1	0.134		1
32	0.089	0.859	1	0.089	0.859	1	0.089	0.860	1
64	0.055	0.951	1	0.055	0.951	1	0.055	0.951	1
128	0.032	0.988	1	0.032	0.988	1	0.032	0.988	1
256	0.018	0.999	1	0.018	0.999	1	0.018	0.999	1

Table 3
Results for example with boundary layers on all sides and orthogonal projection.

$\mathcal{M}_h = R_h^{\text{orth}} BV_h$									
N	$\varepsilon = 1$			$\varepsilon = 10^{-2}$			$\varepsilon = 10^{-4}$		
	Error	Rate	It	Error	Rate	It	Error	Rate	It
16	0.0027		3	0.0177		4	0.073		5
32	0.0008	2.490	3	0.0054	2.509	4	0.038	1.417	8
64	0.0003	2.203	3	0.0018	2.190	4	0.016	1.708	12
128	9.0e−05	2.022	3	0.0005	2.191	5	0.006	1.903	19
256	3.1e−05	1.907	3	0.0002	1.910	5	0.002	1.978	28
N	$\varepsilon = 10^{-8}$			$\varepsilon = 10^{-12}$			$\varepsilon = 10^{-16}$		
	Error	Rate	It	Error	Rate	It	Error	Rate	It
16	0.073		4	0.073		5	0.073		6
32	0.038	1.419	6	0.038	1.419	8	0.038	1.419	10
64	0.016	1.710	9	0.016	1.711	12	0.016	1.711	16
128	0.006	1.903	12	0.006	1.906	19	0.006	1.906	25
256	0.002	1.972	17	0.002	1.981	28	0.002	1.981	40

Table 4
Results for example with boundary layers on all sides and lump projection.

$\mathcal{M}_h = R_h^{\text{lump}} BV_h$									
N	$\varepsilon = 1$			$\varepsilon = 10^{-2}$			$\varepsilon = 10^{-4}$		
	Error	Rate	It	Error	Rate	It	Error	Rate	It
16	0.0048		4	0.0281		5	0.099		3
32	0.0017	2.222	4	0.0088	2.455	6	0.058	1.148	4
64	0.0006	2.042	4	0.0028	2.197	6	0.027	1.515	6
128	0.0002	1.933	4	0.0010	2.052	7	0.010	1.839	8
256	7.3e−05	1.860	4	0.0003	1.898	7	0.003	1.972	11
N	$\varepsilon = 10^{-8}$			$\varepsilon = 10^{-12}$			$\varepsilon = 10^{-16}$		
	Error	Rate	It	Error	Rate	It	Error	Rate	It
16	0.100		4	0.100		5	0.100		7
32	0.058	1.153	7	0.058	1.153	9	0.058	1.154	11
64	0.027	1.524	10	0.027	1.524	13	0.027	1.524	17
128	0.010	1.855	14	0.010	1.856	21	0.010	1.856	27
256	0.003	2.015	21	0.003	2.016	32	0.003	2.016	44

7.3. Example with nonhomogeneous boundary condition

For this example, we solved (1.1) on the unit square with variable coefficient $c = 1 + x^2y^2e^{xy/2}$ and f computed such that the exact solution is

$$\begin{aligned} u(x, y) &= x^3(1 + y^2) + \sin(\pi x^2) + \cos(\pi y/2) \\ &\quad + (x + y)\left(e^{-2x/\sqrt{\varepsilon}} + e^{2(x-1)/\sqrt{\varepsilon}} + e^{-3y/\sqrt{\varepsilon}} + e^{3(y-1)/\sqrt{\varepsilon}}\right), \end{aligned}$$

Table 5
Results for non-homogeneous example, no projection trial space.

$\mathcal{M}_h = BV_h$									
N	$\varepsilon = 1$			$\varepsilon = 10^{-2}$			$\varepsilon = 10^{-4}$		
	Error	Rate	It	Error	Rate	It	Error	Rate	It
16	0.205		1	1.082		1	2.009		1
32	0.103	1.468	1	0.595	1.273	1	1.666	0.398	1
64	0.051	1.355	1	0.306	1.303	1	1.220	0.610	1
128	0.026	1.286	1	0.154	1.273	1	0.791	0.804	1
256	0.013	1.239	1	0.077	1.235	1	0.472	0.921	1
N	$\varepsilon = 10^{-8}$			$\varepsilon = 10^{-12}$			$\varepsilon = 10^{-16}$		
	Error	Rate	It	Error	Rate	It	Error	Rate	It
16	1.989		1	1.988		1	1.988		1
32	1.652	0.394	1	1.652	0.394	1	1.652	0.394	1
64	1.212	0.607	1	1.212	0.607	1	1.212	0.607	1
128	0.786	0.802	1	0.786	0.802	1	0.786	0.802	1
256	0.470	0.920	1	0.470	0.920	1	0.470	0.920	1

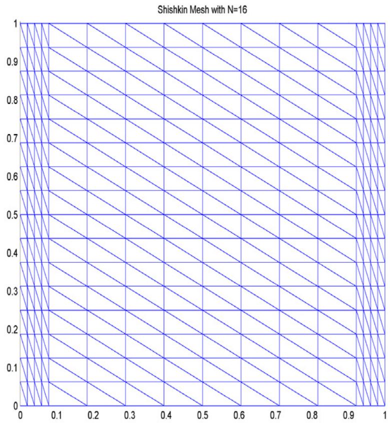
as considered in [6]. The family of Shishkin meshes $\{\mathcal{T}_h\}$ is obtained as in Section 7.2. Table 5 shows results for the no projection trial space and various values of ε. We observe $\mathcal{O}(N^{-1}\ln N)$ convergence and that the order is robust with respect to ε.

7.4. *Example with boundary layers on two sides*

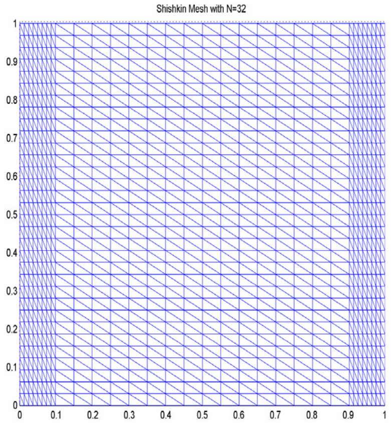
For the last example, we solved (1.1) on the unit square with $c = 2$ and f computed such that the exact solution is given by

$$u(x,y) = y(1-y)\left(1 - e^{-x/\sqrt{\varepsilon}}\right)\left(1 - e^{(x-1)/\sqrt{\varepsilon}}\right),$$

as considered in [11]. Due to the nature of the solution, we expect boundary layers at $x = 0$ and $x = 1$. To this end, we construct the family of Shishkin meshes $\{\mathcal{T}_h\}$ such that the subintervals in the x direction are partitioned as described in Section 6 using $c^* = \sqrt{1/2}$ and $N = 16, 32, 64, 128, 256$, while the partition in the y direction is uniform with N subintervals. The figure below shows the mesh generated with $\varepsilon = 10^{-4}$ and $N = 16, 32$, respectively.



Shishkin mesh, $N = 16$



Shishkin mesh, $N = 32$

Table 6 shows results for the no projection trial space for various values of ε. As in the previous two examples, we observe $\mathcal{O}(N^{-1}\ln N)$ convergence in the balanced norm for the no projection trial space. Tables 7 and 8 display results for the orthogonal and lump projection type spaces, respectively. We observe close to $\mathcal{O}((N^{-1}\ln N)^2)$ convergence in the balanced norm. Furthermore, the order of convergence is robust with respect to ε.

8. Conclusion

We presented a saddle point least squares method with nonconforming trial spaces for discretization of mixed variational formulations for solving the reaction–diffusion equation. We observe that the method performs well even for

Table 6
Results for example with boundary layers on two sides and no projection trial space.

$\mathcal{M}_h = BV_h$									
N	$\varepsilon = 1$			$\varepsilon = 10^{-2}$			$\varepsilon = 10^{-4}$		
	Error	Rate	It	Error	Rate	It	Error	Rate	It
16	0.0094		1	0.040		1	0.091		1
32	0.0047	1.472	1	0.020	1.460	1	0.062	0.839	1
64	0.0024	1.356	1	0.010	1.353	1	0.038	0.935	1
128	0.0012	1.286	1	0.005	1.285	1	0.022	0.978	1
256	0.0006	1.239	1	0.002	1.238	1	0.013	0.993	1
N	$\varepsilon = 10^{-8}$			$\varepsilon = 10^{-12}$			$\varepsilon = 10^{-16}$		
	Error	Rate	It	Error	Rate	It	Error	Rate	It
16	0.091		1	0.091		1	0.091		1
32	0.061	0.835	1	0.061	0.835	1	0.061	0.835	1
64	0.038	0.934	1	0.038	0.934	1	0.038	0.934	1
128	0.022	0.977	1	0.022	0.977	1	0.022	0.977	1
256	0.013	0.993	1	0.013	0.993	1	0.013	0.993	1

Table 7
Results for example with boundary layers on two sides and orthogonal projection.

$\mathcal{M}_h = R_h^{\text{orth}} BV_h$									
N	$\varepsilon = 1$			$\varepsilon = 10^{-2}$			$\varepsilon = 10^{-4}$		
	Error	Rate	It	Error	Rate	It	Error	Rate	It
16	0.0015		2	0.0110		3	0.050		4
32	0.0005	2.378	2	0.0032	2.573	4	0.025	1.469	7
64	0.0002	2.126	2	0.0011	2.149	4	0.010	1.780	12
128	5.6e−05	1.976	2	0.0004	1.982	4	0.004	1.941	19
256	1.9e−05	1.882	2	0.0001	1.884	4	0.001	1.988	29
N	$\varepsilon = 10^{-8}$			$\varepsilon = 10^{-12}$			$\varepsilon = 10^{-16}$		
	Error	Rate	It	Error	Rate	It	Error	Rate	It
16	0.050		3	0.050		4	0.050		5
32	0.025	1.464	6	0.025	1.464	7	0.025	1.464	9
64	0.010	1.777	9	0.010	1.779	12	0.010	1.779	14
128	0.004	1.932	12	0.004	1.942	19	0.004	1.942	23
256	0.001	1.962	17	0.001	1.989	29	0.001	1.990	41

Table 8
Results for example with boundary layers on two sides and lump projection.

$\mathcal{M}_h = R_h^{\text{lump}} BV_h$									
N	$\varepsilon = 1$			$\varepsilon = 10^{-2}$			$\varepsilon = 10^{-4}$		
	Error	Rate	It	Error	Rate	It	Error	Rate	It
16	0.0024		3	0.0161		4	0.068		3
32	0.0008	2.202	3	0.0051	2.442	5	0.038	1.226	4
64	0.0003	2.032	3	0.0017	2.131	5	0.016	1.637	6
128	0.0001	1.928	3	0.0006	1.972	5	0.006	1.900	8
256	3.8e−05	1.857	3	0.0002	1.974	6	0.002	1.911	10
N	$\varepsilon = 10^{-8}$			$\varepsilon = 10^{-12}$			$\varepsilon = 10^{-16}$		
	Error	Rate	It	Error	Rate	It	Error	Rate	It
16	0.067		4	0.067		4	0.067		5
32	0.038	1.231	7	0.038	1.231	8	0.038	1.231	9
64	0.016	1.650	10	0.016	1.650	14	0.016	1.650	15
128	0.006	1.947	15	0.006	1.948	21	0.006	1.948	28
256	0.002	2.028	21	0.002	2.035	33	0.002	2.035	44

$\varepsilon \approx 0$, and we obtain convergence rates of $\mathcal{O}((N^{-1}\ln N)^2)$ using just piecewise linear approximation and the projection type trial spaces. These rates of convergence are similar to those obtained by Lin and Stynes in [6], where a mixed finite element approach was taken involving $H(\text{div}; \Omega)$ conforming spaces. Compared with their approach, our implementation is simpler due to the use of H^1 linear spaces. Also, when using the projection type spaces we obtain close to $\mathcal{O}((N^{-1}\ln N)^2)$ without the need to post-process the solution, which is the approach taken in [11] to obtain higher order convergence for $\varepsilon^{1/4}\nabla u$ in the L^2 norm.

We plan to combine the SPLS discretization method with known adaptive techniques for designing robust iterative solvers for more general first and second order elliptic PDEs that are parameter dependent, including Maxwell equations and linear elasticity systems [27].

We further plan to investigate SPLS multilevel preconditioning techniques, see [2,28,29], that can be considered on Shishkin-type meshes needed for singularly perturbed problems.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Constantin Bacuta: Conceptualization. **Jacob Jacavage:** Conceptualization.

Acknowledgments

We would like to thank the referees for their help and valuable feedback to help improve the original version of the manuscript.

References

[1] Bacuta C, Jacavage J. A non-conforming saddle point least squares approach for an elliptic interface problem. Comput Methods Appl Math 2019;19(3):399–414.
[2] Bacuta C, Jacavage J. Saddle point least squares preconditioning of mixed methods. Comput Math Appl 2019;77(5):1396–407.
[3] Bacuta C, Jacavage J. Least squares preconditioning for mixed methods with nonconforming trial spaces. Appl Anal 2020;1–20, preprint.
[4] Bacuta C, Qirko K. A saddle point least squares approach to mixed methods. Comput Math Appl 2015;70(12):2920–32.
[5] Apel T. Anisotropic finite elements: local estimates and applications. In: Advances in numerical mathematics. Stuttgart: B. G. Teubner; 1999.
[6] Lin R, Stynes M. A balanced finite element method for singularly perturbed reaction–diffusion problems. SIAM J Numer Anal 2012;50(5):2729–43.
[7] Roos HG, Schopf M. Convergence and stability in balanced norms of finite element methods on shishkin meshes for reaction–diffusion problems: Convergence and stability in balanced norms. ZAMM J Appl Math Mech: Z Angew Math Mech 2014;95(6):551–65.
[8] Clavero C, Gracia JL, O'Riordan E. A parameter robust numerical method for a two dimensional reaction–diffusion problem. Math Comp 2005;74:1743–58.
[9] Heuer Norbert, Karkulik Michael. A robust DPG method for singularly perturbed reaction–diffusion problems. SIAM J Numer Anal 2017;55(3):1218–42.
[10] Roos HG, Stynes M, Tobiska L. Robust numerical methods for singularly perturbed differential equations: Convection-diffusion-reaction and flow problems. Springer series in computational mathematics, 2nd ed.. vol. 24, Springer Berlin Heidelberg; 2008.
[11] Li J. Convergence and superconvergence analysis of finite element methods on highly nonuniform anisotropic meshes for singularly perturbed reaction–diffusion problems. Appl Numer Math 2001;36(2):129–54.
[12] Li J, Navon IM. Uniformly convergent finite element methods for singularly perturbed elliptic boundary value problems i: Reaction–diffusion type. Comput Math Appl 1998;35(3):57–70.
[13] Lin R. Discontinuous discretization for least-squares formulation of singularly perturbed reaction–diffusion problems in one and two dimensions. SIAM J Numer Anal 2008;47(1):89–108.
[14] Lin R. Discontinuous galerkin least-squares finite element methods for singularly perturbed reaction–diffusion problems with discontinuous coefficients and boundary singularities. Numer Math 2009;112(2):295–318.
[15] Linß T. Layer-adapted meshes for reaction-convection-diffusion problems. Lecture notes in mathematics, Berlin: Springer-Verlag; 2010.
[16] Aziz A, Babuška I. Survey lectures on mathematical foundations of the finite element method. In: Aziz A, editor. The mathematical foundations of the finite element method with applications to partial differential equations. 1972.
[17] Bacuta C. Schur complements on Hilbert spaces and saddle point systems. J Comput Appl Math 2009;225(2):581–93.
[18] Bacuta C, Monk P. Multilevel discretization of symmetric saddle point systems without the discrete LBB condition. Appl Numer Math 2012;62(6):667–81.
[19] Bacuta C, Qirko K. A saddle point least squares approach for primal mixed formulations of second order PDEs. Comput Math Appl 2017;73(2):173–86.
[20] Cohen A, Dahmen W, Welper G. Adaptivity and variational stabilization for convection–diffusion equations. ESAIM Math Model Numer Anal 2012;46(5):1247–73.
[21] Broersen Dirk, Stevenson Rob. A robust Petrov-Galerkin discretisation of convection–diffusion equations. Comput Math Appl 2014;68(11):1605–18.
[22] Chan J, Heuer N, Bui-Thanh T, Demkowicz L. A robust DPG method for convection-dominated diffusion problems II: adjoint boundary conditions and mesh-dependent test norms. Comput Math Appl 2014;67(4):771–95.
[23] Chan J, Heuer N, Bui-Thanh T, Demkowicz L. A robust DPG method for convection-dominated diffusion problems II: Adjoint boundary conditions and mesh-dependent test norms. Comput Math Appl 2014;67(4):771–95.
[24] Demkowicz L, Gopalakrishnan J. A class of discontinuous Petrov-Galerkin methods. Part I: the transport equation. Comput Methods Appl Mech Engrg 2010;199(23–24):1558–72.
[25] Morton KW, Barrett JW. Optimal Petrov-Galerkin methods through approximate symmetrization. IMA J Numer Anal 1981;1(4):439–68.
[26] Shishkin GI. Grid approximation of singularly perturbed boundary value problems with a regular boundary layer. Sov J Numer Anal Math Model 1989;4(5):397–417.
[27] Bacuta C, Bramble JH. Regularity estimates for solutions of the equations of linear elasticity in convex plane polygonal domains. Z Angew Math Phys (ZAMP) 2003;54:874–8.
[28] Bacuta C, Bramble JH, Pasciak J. New interpolation results and applications to finite element methods for elliptic boundary value problems. East-West J Numer Math 2001;9(3):179–98.
[29] Xu J. Iterative methods by space decomposition and subspace correction. SIAM Rev 1992;34(4):581–613.