

EFFICIENT DISCRETIZATION AND PRECONDITIONING OF THE SINGULARLY PERTURBED REACTION-DIFFUSION PROBLEM

CONSTANTIN BACUTA, DANIEL HAYES, AND JACOB JACAVAGE

ABSTRACT. We consider the reaction diffusion problem and present efficient ways to discretize and precondition in the singular perturbed case when the reaction term dominates the equation. Using the concepts of optimal test norm and saddle point reformulation, we provide efficient discretization processes for uniform and non-uniform meshes. We present a preconditioning strategy that works for a large range of the perturbation parameter. Numerical examples to illustrate the efficiency of the method are included for a problem on the unit square.

1. INTRODUCTION

We consider the singularly perturbed reaction diffusion problem

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

for non-negative constant ε and $c_{max} \geq c(x) \geq c_{min} > 0$ on Ω , a bounded domain in \mathbb{R}^d . We focus on the reaction dominated case, i.e., $\varepsilon \ll 1$. The discretization of the equation arises in solving practical PDE models, such as heat transfer problems in thin domains [2] as well as when using small step sizes in implicit time discretizations of parabolic reaction diffusion type problems [29]. The solutions to these problems are characterized by exponential boundary layers which pose numerical challenges due to the ε -dependence of the error estimates and of the stability constants.

The standard variational formulation for (1.1) is: Find $u \in H_0^1(\Omega)$ such that

$$(1.2) \quad \varepsilon(\nabla u, \nabla v) + (cu, v) = (f, v) \quad \text{for all } v \in H_0^1(\Omega).$$

To discretize (1.2), we propose a mixed variational formulation that allows for higher order of approximation and for efficient preconditioning. Towards efficient preconditioning, we introduce a simplified version of the Bramble, Pasciak, and Vasilevski (BVP) multilevel preconditioner [14], that works for nested spaces corresponding to uniform refinement meshes.

2000 *Mathematics Subject Classification.* 74S05, 74B05, 65N22, 65N55.

Key words and phrases. least squares, saddle point systems, mixed methods, optimal test norm, Uzawa conjugate gradient, preconditioning.

The work was supported by NSF-DMS 2011615.

A standard discretization of (1.2) can be solved by using a PCG algorithm on $V_h \subset H_0^1(\Omega)$ with the BVP or the simplified preconditioner we introduce. The process is efficient for a certain range of $\varepsilon \geq \varepsilon_0$. To improve the rate of convergence (in the energy norm) for the case $\varepsilon < \varepsilon_0$, we proceed by using two main tools. The first is a mixed formulation that is suitable to Saddle Point Least Squares (SPLS) discretization [7, 8, 9, 11]. This allows for a higher order approximation of the flux ∇u using the same discretization spaces as in the standard discretization of (1.1). The second tool is based on using refined meshes towards the boundary layer(s) of the problem. Even though the refinement approach presented in this paper is focused on the case in which Ω is the unit square in \mathbb{R}^2 , as well as the use of Shiskin meshes, the ideas of mixed reformulation and multilevel preconditioning extend to more general cases of $\Omega \subset \mathbb{R}^d$.

Classical first-order system least squares methods for the reaction diffusion problem can be found in [1, 18, 29]. However, our approach of saddle point least squares reformulation is different. We reformulate the standard variational formulation (1.2) as a mixed variational formulation first, and then further reformulate it as a saddle point problem that allows preconditioning and extra approximability of the flux by using standard finite element spaces. The approach is related to the Lagrange multiplier approach that leads to Stokes type systems. In our case, the Lagrange multiplier is the variable of interest. This idea was used before in other particular problems, see e.g., [10, 20, 23, 25].

Many aspects of the SPLS formulation are also common to the DPG approach [13, 17, 21, 22, 24, 27]. The SPLS method as a general method for solving mixed variational formulations is presented in [7, 8, 9, 11]. In this paper, we also combine SPLS formulation with the concept of optimal test norm [16, 19, 20, 21, 24, 25, 27] and a general preconditioning technique, introduced in [7], in order to improve the stability and approximability of the final discretization process.

The paper is organized as follows. In Section 2, we review the Bramble-Pasciak-Vasilevski multilevel preconditioning technique and introduce a simplified preconditioner for the reaction diffusion equation. In Section 3, we introduce the notation and review the SPLS formulation, discretization, preconditioning, and the concept of optimal test norms. Sections 4 and 5 detail how to apply the general SPLS theory to the reaction diffusion problem. Numerical results are included in Section 6.

2. SPECTRAL REPRESENTATION OF H^1 NORM AND MULTILEVEL PRECONDITIONING

2.1. Abstract formulation. Assume V is a Hilbert space with inner product $a(\cdot, \cdot)$ and that we have nested spaces $V_0 \subset V_1 \subset \dots \subset V_J = V_h \subset V$. We let (\cdot, \cdot) be another inner product on V that induces a weaker (than $a(\cdot, \cdot)$) norm on V , and denote by $Q_j : V \rightarrow V_j$ the orthogonal projection with

respect to the (\cdot, \cdot) inner product for $j = 0, 1, \dots, J$. We will further assume that the following norm equivalence

$$(2.1) \quad \sum_{k=0}^J \lambda_k \|(Q_k - Q_{k-1})v\|^2 \simeq a(v, v),$$

holds on the entire space $V_h = V_J$. Here, the λ_j 's are positive constants, $Q_{-1} = 0$, and Q_J coincides with the identity on V_J . Furthermore, we assume the equivalence constants associated with the symbol " \simeq " are independent of h or J .

Let $c(\cdot, \cdot)$ be another symmetric bilinear form defined on V and define the operators $A_h : V_h \rightarrow V_h$ and $C_h : V_h \rightarrow V_h$ by

$$(A_h u, v) = a(u, v), \quad (C_h u, v) = c(u, v) \quad \text{for all } v \in V_h.$$

We assume that there are positive constants μ_1, \dots, μ_J such that

$$(2.2) \quad \sum_{k=0}^J \mu_k \|(Q_k - Q_{k-1})v\|^2 \simeq c(v, v),$$

with equivalent constants independent of h or J . The goal of this section is to present a simple multilevel preconditioning technique for the following problem: Find $u_h \in V_h$ such that

$$\tilde{a}(u_h, v) := \epsilon a(u_h, v) + c(u_h, v) = (f, v) \quad \text{for all } v \in V_h,$$

or in operator form

$$(\epsilon A_h + C_h)u_h = f_h := Q_h f.$$

2.2. Bramble-Pasciak-Vassilevski multilevel preconditioner. Using that

$$\begin{aligned} Q_k Q_j &= Q_j Q_k = Q_j, \quad j \leq k \\ (Q_k - Q_{k-1})(Q_j - Q_{j-1}) &= (Q_j - Q_{j-1})\delta_{j,k}, \end{aligned}$$

the equivalences (2.1) and (2.2) can be written in the operator forms

$$\begin{aligned} L_h &:= \sum_{j=0}^J \lambda_j (Q_j - Q_{j-1}) \simeq A_J = A_h, \\ K_h &:= \sum_{j=0}^J \mu_j (Q_j - Q_{j-1}) \simeq C_J = C_h, \end{aligned}$$

respectively. Using the above (projection type) spectral representations and the properties of the projections, we obtain that

$$\begin{aligned} \epsilon L_h + C_h &= \sum_{j=0}^J (\epsilon \lambda_j + \mu_j) (Q_j - Q_{j-1}) \simeq \epsilon A_h + C_h, \\ P_h &:= \sum_{j=0}^J (\epsilon \lambda_j + \mu_j)^{-1} (Q_j - Q_{j-1}) = (\epsilon L_h + C_h)^{-1}, \end{aligned}$$

are symmetric positive definite (discrete) operators on V_h . In addition, P_h is a (uniform) preconditioner for $\epsilon A_h + C_h$.

The BPV approach further modifies P_h by using another (sum of local projections type) operator that avoids computing the action of Q_k (hence mass matrix inversion). In the BPV approach, the projections $Q_k - Q_{k-1}$ are replaced by $(\tilde{Q}_k - \tilde{Q}_{k-1})^2$, where \tilde{Q}_k is given by

$$(2.3) \quad \tilde{Q}_k v_h := \sum_{i=1}^{n_k} \frac{(v_h, \varphi_i^k)}{(1, \varphi_i^k)} \varphi_i^k \quad \text{for all } v_h \in V_h,$$

where $\{\varphi_i^k, i = 1, 2, \dots, n_k\}$ is a basis for V_k .

2.3. A simplified BVP (sBVP) preconditioner. A more general form of the preconditioner P_h is given by

$$P_h = \sum_{j=0}^J \gamma_j (Q_j - Q_{j-1}),$$

where γ_j is a positive real number for each j . We further note that by using “summation by parts” for the general form of P_h , we obtain

$$\begin{aligned} P_h f_h &= \sum_{k=0}^J \gamma_k (Q_k - Q_{k-1}) f_h \\ &= \left(\sum_{k=0}^J \gamma_k Q_k - \sum_{k=1}^J \gamma_k Q_{k-1} \right) f_h \\ &= \left(\sum_{k=0}^J \gamma_k Q_k - \sum_{k=0}^{J-1} \gamma_{k+1} Q_k \right) f_h \\ &= \gamma_J f_h + \sum_{k=0}^{J-1} (\gamma_k - \gamma_{k+1}) Q_k f_h. \end{aligned}$$

We are able to further simplify the preconditioner P_h under three assumptions that are easily satisfied in our applications. We first assume that $\{\gamma_j\}$ is a decreasing sequence of positive numbers:

$$(A1) \quad \gamma_0 \geq \gamma_1 \geq \dots \geq \gamma_J > 0.$$

Second, we let $\{\tilde{Q}_k : V_h \rightarrow V_k\}$, with $k = 0, 1, \dots, J-1$, be any family of linear operators satisfying

$$(A2) \quad (\tilde{Q}_k v_h, w_h) = (v_h, \tilde{Q}_k w_h) \quad \text{for all } v_h, w_h \in V_h,$$

and

$$(A3) \quad c_1(Q_k v_h, v_h) \leq (\tilde{Q}_k v_h, v_h) \leq c_2(Q_k v_h, v_h) \quad \text{for all } v_h \in V_h,$$

where c_1, c_2 are positive constants independent of h . We define a new operator $\tilde{P}_h : V_h \rightarrow V_h$ by

$$\tilde{P}_h f_h = \lambda_J f_h + \sum_{k=0}^{J-1} (\gamma_k - \gamma_{k+1}) \tilde{Q}_k f_h.$$

Lemma 2.1. *Under the assumptions (A1), (A2) and (A3) we have that*

$$(2.4) \quad c_1(P_h v_h, v_h) \leq (\tilde{P}_h v_h, v_h) \leq c_2(P_h v_h, v_h) \quad \text{for all } v_h \in V_h.$$

Consequently, if P_h is a uniform preconditioner for A_h , then \tilde{P}_h is also a uniform preconditioner for A_h .

Proof. First, we note that due to (A1) and (A2) we have that \tilde{P}_h is a symmetric positive definite operator on V_h . Multiplying the inequalities in (A3) by the appropriate positive scalars and summing up the new inequalities, we obtain (2.4). \square

2.4. Preconditioning the reaction diffusion problem. The standard variational formulation of problem (1.1) is: Find $u \in H_0^1(\Omega)$ such that

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

where (\cdot, \cdot) denotes the standard L^2 inner product for scalar or vector functions. We consider the nested sequence of spaces $\{V_k\}_{k=0, \overline{J}}$ of continuous piecewise linear functions associated with the uniformly refined meshes $\{\mathcal{T}_k\}_{k=0, \overline{J}}$ on Ω , and define the operators

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

$$(C_h u, v) = (cu, v) \quad \text{for all } u, v \in V_h.$$

That is, A_h is the discrete Laplacian operator and C_h is the discretization of the multiplication by the function $c(x)$ operator. It is well known, see [14, 15], that for $\lambda_j = 1/h_j^2 \approx 4^j$ we have

$$L_h := \sum_{i=0}^J \lambda_j (Q_j - Q_{j-1}) \simeq A_J = A_h.$$

Using that $\sum_{j=0}^J (Q_j - Q_{j-1}) = I$, for $\mu_j := c^*$, where c^* is any value in $[c_{min}, c_{max}]$, we obtain

$$K_h := \sum_{j=0}^J \mu_j (Q_j - Q_{j-1}) \simeq (c^* I) \simeq C_h.$$

Thus, the BPV preconditioner for $\epsilon A_h + C_h$ becomes

$$(2.5) \quad \begin{aligned} B_J^{BVP} &:= \sum_{j=0}^J (\epsilon/h_j^2 + c^*)^{-1} (Q_j - Q_{j-1}) \\ &= \gamma_J I + \sum_{j=0}^{J-1} (\gamma_j - \gamma_{j+1}) Q_j, \end{aligned}$$

where $\gamma_j = (\epsilon/h_j^2 + c^*)^{-1}$.

It was proved in [14] that B_J^{BVP} is a uniform preconditioner for $\epsilon A_h + C_h$. Using the family $\{\tilde{Q}_k : V_h \rightarrow V_k\}_{k=0, \overline{J-1}}$ defined in (2.3), with $\{\varphi_i^k, i = 1, 2, \dots, n_k\}$ the standard nodal basis on V_k , we can define the simplified BPV (sBVP) preconditioner for the reaction diffusion problem

$$(2.6) \quad B_J^{sBVP} := \gamma_J I + \sum_{j=0}^{J-1} (\gamma_j - \gamma_{j+1}) \tilde{Q}_j,$$

where $\gamma_j = (\epsilon/h_j^2 + c^*)^{-1}$. To prove that B_J^{sBVP} is a uniform preconditioner for $\epsilon A_h + C_h$, we only need to check that (A1) – (A3) are satisfied. Assumption (A1) is satisfied as the function $h \rightarrow (\epsilon/h^2 + c^*)^{-1}$ is decreasing on $(0, 1]$. From the definition of \tilde{Q}_k in (2.3), one can easily verify (A2). We will prove that (A3) holds next.

Lemma 2.2. *Let $\{V_k\}_{k=0, \overline{J}}$ be a nested sequence of spaces of continuous piecewise linear functions associated with a set of quasi-uniform meshes $\{\mathcal{T}_k\}_{k=0, \overline{J}}$ on Ω , and assume that \tilde{Q}_k is defined as in (2.3). Then assumption (A3) holds.*

Proof. Define the diagonal matrix D with entries $D_{ii} = (1, \varphi_i^k)$, and let M be the mass matrix for the basis $\{\varphi_i^k, i = 1 : n_k\}$. Using that the mesh \mathcal{T}_k is quasi-uniform (could be just locally), we obtain

$$(v_h, v_h) \simeq \sum_{i=1}^{n_k} (1, \varphi_i) v_h^2(z_i) \quad \text{for all } v_h \in V_h,$$

where the uniformity constants are independent of k . Here, z_i corresponds to the nodal function φ_i^k , i.e., $\varphi_i^k(z_j) = \delta_{ij}$. This is equivalent to

$$(2.7) \quad M \simeq D \text{ or } D^{-1} \simeq M^{-1},$$

which implies

$$MD^{-1}M \simeq M.$$

From (2.3), we obtain

$$(\tilde{Q}_k v_h, v_h) = \sum_{i=1}^{n_k} \frac{(v_h, \varphi_i^k)^2}{(1, \varphi_i^k)} = \sum_{i=1}^{n_k} \frac{(Q_k v_h, \varphi_i^k)^2}{(1, \varphi_i^k)}.$$

Let $Q_k v_h = \sum_{j=1}^{n_k} \alpha_j^k \varphi_j^k$ and $\alpha = (\alpha_j^k)_{j=1: n_k}$. Then, using $MD^{-1}M \simeq M$,

$$(\tilde{Q}_k v_h, v_h) = \sum_{i=1}^{n_k} \frac{(\sum_{j=1}^{n_k} \alpha_j^k \varphi_j^k, \varphi_i^k)^2}{(1, \varphi_i^k)} = (D^{-1}M\alpha, M\alpha)_e \simeq (M\alpha, \alpha)_e,$$

where $(\cdot, \cdot)_e$ is the Euclidian inner product. Since

$$(M\alpha, \alpha)_e = (Q_k v_h, Q_k v_h) = (Q_k v_h, v_h),$$

assumption (A3) holds. \square

Remarks on the implementation of B_J^{sBVP} are included in the Appendix. Other related works on preconditioning singularly perturbed reaction-diffusion problem can be found in [26] for the finite element discretization on quasi-uniform meshes, where a more general Additive Schwartz preconditioner is proposed, and in [30] for finite difference discretization on fitted meshes, where a block-structured preconditioning approach is proposed.

3. THE NOTATION AND THE GENERAL SPLS APPROACH

We now review the main ideas and concepts for the SPLS formulation, discretization, and preconditioning of a general mixed variational formulation. Let V and Q be Hilbert spaces and $F \in V^*$. We are interested in problems of the form: Find $p \in Q$ such that

$$(3.1) \quad b(v, p) = \langle F, v \rangle \quad \text{for all } v \in V,$$

where $b(\cdot, \cdot)$ is a continuous bilinear form on $V \times Q$ satisfying the following inf – sup condition on $V \times Q$:

$$(3.2) \quad \inf_{p \in Q} \sup_{v \in V} \frac{b(v, p)}{|v| \|p\|} = m > 0.$$

We view Q , the trial space in (3.1), 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 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 further assume that $b(\cdot, \cdot)$ has a continuous extension to a bilinear form on $V \times \tilde{Q}$ satisfying

$$(3.3) \quad \sup_{p \in \tilde{Q}} \sup_{v \in V} \frac{b(v, p)}{|v| \|p\|} = M < \infty.$$

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

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

In this paper, we assume that $V_0 := \text{Ker}(B) = \{0\}$. Most of the considerations in this section extend to a nontrivial kernel V_0 , see [9]. It is well known that if a bounded form $b : V \times Q \rightarrow \mathbb{R}$ satisfies (3.2), then problem (3.1) has a unique solution, see e.g., [3, 4]. The standard *saddle point reformulation* of (3.1) (see [10, 11, 12, 20]) is: Find $(w, p) \in (V, Q)$ such that

$$(3.4) \quad \begin{aligned} a_0(w, v) + b(v, p) &= \langle F, v \rangle & \text{for all } v \in V, \\ b(w, q) &= 0 & \text{for all } q \in Q. \end{aligned}$$

3.1. The concept of optimal test norm. If we assume that $\text{Range}(B) \subset Q$ and that the operator $B : V \rightarrow Q$ is injective ($V_0 = \text{Ker}(B) = \{0\}$) then, as in [16, 19, 20, 21, 25], we can define an equivalent norm on V

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

that is operator dependent. We will refer to this as the *optimal test norm*. By replacing the form $a_0(\cdot, \cdot)$ in (3.4) with the inner product induced by the *optimal test norm*, i.e., $a_{opt}(u, v) := (Bu, Bv)_Q$, we obtain that both the continuity constant M and the inf – sup constant m are equal to 1. Thus, the stability (at the continuous level) of the new saddle point formulation is optimal.

3.2. The abstract variational formulation at the discrete level. The non-conforming (trial space) *SPLS discretization* of (3.1) is defined as a saddle point discretization of (3.4) 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 standard discrete sup – sup and inf – sup conditions hold for the pair (V_h, \mathcal{M}_h) with constants M_h and m_h respectively. The discrete mixed variational formulation of (3.1) is: Find $p_h \in \mathcal{M}_h$ such that

$$(3.5) \quad b(v_h, p_h) = \langle F, v_h \rangle \quad \text{for all } v_h \in V_h.$$

In general, this might not have a unique solution. However, discretization of (3.4): Find $(w_h, p_h) \in V_h \times \mathcal{M}_h$ such that

$$(3.6) \quad \begin{aligned} a_0(w_h, v_h) + b(v_h, p_h) &= \langle F, v_h \rangle && \text{for all } v_h \in V_h, \\ b(w_h, q_h) &= 0 && \text{for all } q_h \in \mathcal{M}_h, \end{aligned}$$

always has a unique solution. The variational formulation (3.6) is the *non-conforming saddle point least squares* (n-c SPLS) *discretization* of (3.1).

Using (\cdot, \cdot) , another (weaker) inner product on V , we can define the discrete operator $A_h : V_h \rightarrow V_h$ associated with the form $a_0(\cdot, \cdot)$ on V_h by

$$(A_h v_h, w_h) = a_0(v_h, w_h) \quad \text{for all } v_h, w_h \in V_h$$

and the linear operators $B_h : V_h \rightarrow \mathcal{M}_h$ and $B_h^* : \mathcal{M}_h \rightarrow V_h$ by

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

The Schur complement of (3.6) is denoted by $S_h = B_h A_h^{-1} B_h^*$. 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 $p_h \in \mathcal{M}_h \subset \tilde{Q}$.

3.3. Constructing a discrete trial space from a general test space.

Let V_h be a *finite element subspace* of V . Following [8, 9], we provide a general construction of discrete trial spaces \mathcal{M}_h , defined using the operator B associated with the original problem (3.1). 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 $R_h : \tilde{Q} \rightarrow \tilde{\mathcal{M}}_h$ by

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

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

$$(3.7) \quad k_1 \|q_h\| \leq \|q_h\|_h \leq k_2 \|q_h\| \quad \text{for all } q_h \in \tilde{\mathcal{M}}_h.$$

Using the operator R_h , we define \mathcal{M}_h as

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

The following proposition, see [4] and [9], gives a sufficient condition on R_h to ensure that the family $\{(V_h, R_h B V_h)\}$ is stable.

Proposition 3.1. *Assume that*

$$(3.8) \quad \|R_h q_h\|_h \geq \tilde{c} \|q_h\| \quad \text{for all } q_h \in B V_h,$$

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

$$(3.9) \quad \inf_{p_h \in \mathcal{M}_h} \sup_{v_h \in V_h} \frac{b(v_h, p_h)}{|v_h| \|p_h\|_h} \geq \tilde{c} m > 0,$$

where m is defined in (3.2).

As a consequence of Proposition 3.1, we have that (3.6) has a unique solution $(w_h, p_h) \in (V_h, \mathcal{M}_h)$ and $w_h = 0$. Regarding the *approximability property* of the projection type trial space, the following proposition was proved in [9].

Proposition 3.2. *If p is the solution of (3.1) and p_h is the second component of the solution of (3.6), then*

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

with $C = 1 + \frac{1}{k_1 \tilde{c}}$, where k_1 and \tilde{c} were introduced in (3.7) and (3.8), respectively.

Remark 3.3. *The choice of $\tilde{\mathcal{M}}_h$ is important. In practice V_h is a space of continuous piecewise polynomials, and by applying the (differentiation) operator B we obtain discontinuous functions. The representation operator $R_h : \tilde{Q} \rightarrow \tilde{\mathcal{M}}_h$ acting on $B V_h$ can be viewed also as a smoothing operator, as the range $\tilde{\mathcal{M}}_h$ is a subspace of \tilde{Q} consisting of continuous functions. The SPLS discretization with $\tilde{\mathcal{M}}_h$ as trial space can be viewed as a (smoothing type) recovery or post-processing process.*

3.4. An Uzawa CG iterative solver. Note that a global linear system may be difficult to assemble when solving (3.6) on $(V_h, \mathcal{M}_h = R_h B V_h)$, especially if the operator R_h involves a global projection. In this case, bases for the trial spaces \mathcal{M}_h might be difficult to find. One can solve (3.6) and avoid building a basis for \mathcal{M}_h by using an Uzawa Conjugate Gradient (UCG) algorithm.

Algorithm 3.4. (*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

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

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

$$\text{(UCG1)} \quad a_0(h_j, v_h) = -b(v_h, d_j) \quad \text{for all } v_h \in V_h$$

$$\text{(UCG}\alpha) \quad \alpha_j = -\frac{(q_j, q_j)_h}{b(h_j, q_j)}$$

$$\text{(UCG2)} \quad p_j = p_{j-1} + \alpha_j d_j$$

$$\text{(UCG3)} \quad w_{j+1} = w_j + \alpha_j h_j$$

$$\text{(UCG4)} \quad (q_{j+1}, q)_h = b(w_{j+1}, q) \quad \text{for all } q \in \mathcal{M}_h$$

$$\text{(UCG}\beta) \quad \beta_j = \frac{(q_{j+1}, q_{j+1})_h}{(q_j, q_j)_h}$$

$$\text{(UCG6)} \quad d_{j+1} = q_{j+1} + \beta_j d_j.$$

Remark 3.5. From **(UCG4)**, we have that $q_{j+1} = B_h w_{j+1}$. Thus, 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. 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 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

$$(3.10) \quad w_1 = A_h^{-1}(f_h - B_h^* p_0), \quad \text{and} \quad h_j = -A_h^{-1}(B_h^* d_j).$$

In order to build an efficient solver for (3.1), we modify Algorithm 3.4 by replacing the action of A_h^{-1} with the action of a suitable preconditioner.

3.5. Preconditioning the SPLS discretization. We summarize a general preconditioning framework to approximate the solution of (3.1) that is presented in [5, 7]. We plan to combine this framework with the new concept of *optimal* test norm. Let $P_h : V_h \rightarrow V_h$ be a general (preconditioning) operator that is equivalent to A_h^{-1} in the sense that

$$(3.11) \quad (P_h f_h, g_h) = (f_h P_h g_h) \quad \text{for all } f_h, g_h \in V_h,$$

and

$$(3.12) \quad m_1^2 |v_h|^2 \leq a_0(P_h A_h v_h, v_h) \leq m_2^2 |v_h|^2,$$

where the positive constants m_1^2, m_2^2 are the smallest and largest eigenvalues of $P_h A_h$, respectively. Condition (3.12) is equivalent with the fact that the condition number of $P_h A_h$ satisfies

$$(3.13) \quad \kappa(P_h A_h) = \frac{m_2^2}{m_1^2}.$$

It was proved in [5, 7] that the saddle point discretization (3.6) would not lose stability and approximability properties if $\kappa(P_h A_h)$ is independent of h and A_h^{-1} is replaced by P_h . The replacement of the action of A_h^{-1} leads to solving

$$(3.14) \quad \begin{aligned} w_h &= P_h(f_h - B_h^* p_h), \\ B_h w_h &= 0. \end{aligned}$$

The Schur complement associated with the modified problem (3.14) is

$$\tilde{S}_h = B_h P_h B_h^*.$$

The corresponding version of Algorithm 3.4 to solve (3.14) is the following Uzawa Preconditioned Conjugate Gradient (UPCG) algorithm.

Algorithm 3.6. (*UPCG*) *Algorithm for Mixed Methods*

Step 1: Choose any $p_0 \in \mathcal{M}_h$. Compute $u_1 \in V_h, q_1, d_1 \in \mathcal{M}_h$ by

$$\begin{aligned} u_1 &= P_h(F_h - B_h^* p_0) \\ q_1 &= B_h u_1, \quad d_1 := q_1. \end{aligned}$$

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

$$\begin{aligned} \text{(PCG1)} \quad & h_j = -P_h(B_h^* d_j) \\ \text{(PCG}\alpha) \quad & \alpha_j = -\frac{(q_j, q_j)_Q}{b(h_j, q_j)} \\ \text{(PCG2)} \quad & p_j = p_{j-1} + \alpha_j d_j \\ \text{(PCG3)} \quad & u_{j+1} = u_j + \alpha_j h_j \\ \text{(PCG4)} \quad & q_{j+1} = B_h u_{j+1}, \\ \text{(PCG}\beta) \quad & \beta_j = \frac{(q_{j+1}, q_{j+1})_Q}{(q_j, q_j)_Q} \\ \text{(PCG6)} \quad & d_{j+1} = q_{j+1} + \beta_j d_j. \end{aligned}$$

According to [7], the UPCG iterations p_j converge to the solution p_h of (3.14), and the rate of convergence for $\|p_j - p_h\|_{\tilde{S}_h}$ depends on the condition number of \tilde{S}_h that satisfies

$$(3.15) \quad \kappa(\tilde{S}_h) \leq \kappa(S_h) \cdot \kappa(P_h A_h).$$

4. SPLS FORMULATION FOR THE REACTION DIFFUSION PROBLEMS

In this section, we consider the SPLS discretization for the reaction diffusion problem using an optimal test norm. The goal is to emphasize how stability and approximability for the mixed formulation can be gained by using the SPLS approach and provide a way of choosing appropriate preconditioners to find an efficient iterative solver. In what follows, (\cdot, \cdot) and $\|\cdot\|$ will denote the standard L^2 inner product and norm, respectively.

To place equation (1.1) into the general SPLS framework, we define the spaces $V := H_0^1(\Omega)$, $\tilde{Q} := L^2(\Omega) \times L^2(\Omega)^d$, and Q as the graph of the operator $\varepsilon\nabla : H_0^1(\Omega) \rightarrow L^2(\Omega)^d$, 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 $b : V \times \tilde{Q} \rightarrow \mathbb{R}$ as

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

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 (1.2) is: Find $\mathbf{p} = \begin{pmatrix} u \\ \varepsilon\nabla u \end{pmatrix} \in Q$ such that

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

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

$$(4.1) \quad \left(\begin{pmatrix} q \\ \mathbf{q} \end{pmatrix}, \begin{pmatrix} p \\ \mathbf{p} \end{pmatrix} \right)_{\tilde{Q}} = (cq, p) + (\varepsilon^{-1}\mathbf{q}, \mathbf{p}) := (q, p)_c + (\mathbf{q}, \mathbf{p})_{\varepsilon^{-1}},$$

which gives us the corresponding norm

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

The operator $B : V \rightarrow \tilde{Q}$ is then given by

$$Bv = \begin{pmatrix} v \\ \varepsilon\nabla v \end{pmatrix} \quad \text{for all } v \in V.$$

We note that

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

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

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

In addition, according to Section 3.1, we have that the continuity constant M of the bilinear form b and the inf – sup constant m are equal to 1. One can also directly check that

$$(4.2) \quad \sup_{v \in V} \frac{b(v, (\varepsilon \nabla u))}{|v|_{opt}} = \|(\varepsilon \nabla u)\|_Q,$$

for any $(\varepsilon \nabla u) \in Q$. While this does lead to optimal continuity and inf – sup constants, inverting the operator associated with $|\cdot|_{opt}$ coincides with solving the original problem. Fortunately, at the discrete level we can replace the action of the discrete operator corresponding to $a_{opt}(\cdot, \cdot)$ by a preconditioner, as presented in Section 2.

5. SPLS DISCRETIZATION FOR THE REACTION DIFFUSION PROBLEMS

In this section, we outline choices for the test and trial spaces involved in SPLS discretization. We take $V_h \subset V = H_0^1(\Omega)$ to be the space of continuous piecewise linear polynomials with respect to the mesh \mathcal{T}_h vanishing on the boundary of Ω . To construct a trial space \mathcal{M}_h , as discussed in Section 3.3, we first define $\tilde{\mathcal{M}}_h \subset \tilde{Q} = L^2(\Omega) \times L^2(\Omega)^d$ as

$$\tilde{\mathcal{M}}_h := M_h \times \varepsilon \mathbf{M}_h,$$

where M_h consists of continuous piecewise linear polynomials with respect to the mesh \mathcal{T}_h (with no restrictions on the boundary) and \mathbf{M}_h is the vector-valued product space in which each component consists of continuous piecewise linear polynomials. Let $\{\phi_1, \dots, \phi_N\}$ denote a nodal basis for M_h with respect to the mesh \mathcal{T}_h and $\{\Phi_1, \dots, \Phi_{2N}\}$ denote a nodal basis for \mathbf{M}_h , where $\Phi_j = (\phi_j, 0)^T$ and $\Phi_{N+j} = (0, \phi_j)^T$ for $j = 1, \dots, N$. Two different choices the projection type trial space, based on the inner product chosen for $\tilde{\mathcal{M}}_h$, are considered. More details about the construction of these trial spaces can be found in [6].

For the first choice of projection trial space, we equip $\tilde{\mathcal{M}}_h$ with the inner product induced from \tilde{Q} . In this case, R_h is the orthogonal projection from \tilde{Q} onto $\tilde{\mathcal{M}}_h$, and we define the trial space as

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

where we use the notation R_h^{orth} to signify equipping $\tilde{\mathcal{M}}_h$ with the induced inner product from \tilde{Q} .

For the second choice of projection trial space, we equip $\tilde{\mathcal{M}}_h$ with an inner product related to lumping the mass matrix. More precisely, for two elements $(\mathbf{q}_h^h), (\mathbf{p}_h^h) \in \tilde{\mathcal{M}}_h$, where

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

we define

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

In this case, the action of $R_h : \tilde{Q} \rightarrow \tilde{\mathcal{M}}_h$ is given by

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

where $Q_h : L^2(\Omega) \rightarrow M_h$ is the orthogonal projection with respect to the weighted inner product $(\cdot, \cdot)_c$ and

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

We then define the trial space as

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

where we use the notation R_h^{lump} to signify equipping $\tilde{\mathcal{M}}_h$ with this type of inner product. For the remainder of this section, we will write R_h without the superscript for simplicity as the results hold for either choice of inner product.

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

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

and the corresponding SPLS discretization, with *optimal test norm*, is: Find $(w_h, \mathbf{p}_h = R_h A \nabla u_h) \in V_h \times \mathcal{M}_h$ such that

$$(5.1) \quad \begin{aligned} a_{\text{opt}}(w_h, v_h) + b(v_h, \mathbf{p}_h) &= (f, v_h) & \text{for all } v_h \in V_h, \\ R_h B w_h &= \mathbf{0}. \end{aligned}$$

In the above problem, as mentioned we can replace the action of the discrete operator corresponding to $a_{\text{opt}}(\cdot, \cdot)$ by a preconditioner. The following result regarding the stability of these spaces was proved in [6].

Theorem 5.1. *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 $\{T_h\}$ that vanish on $\partial\Omega$ and $\mathcal{M}_h = R_h B V_h$. Then the family of spaces $\{(V_h, \mathcal{M}_h)\}$ is stable.*

Remark 5.2. *While the above result assumes $\Omega \subset \mathbb{R}^2$ is a polygonal domain, the result can be extended to polyhedral domains in \mathbb{R}^3 .*

6. NUMERICAL RESULTS

We considered equation (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 [28]. For this problem, the solution has boundary layers on all sides of the unit square. The test space $V_h = \subset H_0^1(\Omega)$ was chosen to be the space of continuous piecewise linear polynomials with respect to the mesh \mathcal{T}_h and $\mathcal{M}_h = R_h B V_h$ as described in Section 5. We used Algorithm 3.6 to solve (5.1) with two types of preconditioners: the sBPV preconditioner introduced in Section 2.3 and a multigrid preconditioner with Gauss-Seidel smoother. We applied the method on both a uniform mesh as well as a Shishkin type mesh, introduced in [32]. Details on the implementation of the sBPV preconditioner can be found in the Appendix.

For the Shishkin mesh, we followed the construction outlined in [31]. We add it here for completion. We first assume the parameter N is an integer multiple of 8. This refers 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

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

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.

In the case of the Shishkin mesh, we also measured the SPLS error in a balanced norm instead of the norm on Q . This is due to the fact that for small ε the L^2 part of the norm on Q dominates, leading to an unbalanced norm not adequate to accurately measure the error, see [29, 31]. More precisely, in this case, we compute

$$Error := \left(\|u - u_h\|^2 + \varepsilon^{1/2} \|\nabla u - R_h \nabla u_h\|^2 \right)^{1/2}.$$

In the Shishkin mesh case, we used a stopping criteria of $\|q_j\|_Q \leq 10^{-10}$ for $10^{-8} \leq \varepsilon \leq 10^{-4}$, and a stopping criteria of $\|q_j\|_Q \leq 10^{-16}$ for $10^{-14} \leq \varepsilon \leq 10^{-10}$. In the case of a uniform mesh, we used a stopping criteria of $\|q_j\|_Q \leq 10^{-8}$ for $\varepsilon \geq 10^{-6}$.

Table 1 displays results using uniform meshes, $\mathcal{M}_h = R_h^{\text{orth}}AV_h$, and both the sBVP and multigrid preconditioners. For $\varepsilon \leq 10^{-3}$ it is shown that the sBVP preconditioner retains a lower iteration count across all levels. The loss of order for small ε in the uniform mesh case is due to the boundary layers and would need further refinements to resolve them.

Lev/ ε	10^{-1}				10^{-2}				10^{-3}			
	Error	Order	(a)	(b)	Error	Order	(a)	(b)	Error	Order	(a)	(b)
1	0.0511	-	6	6	0.0866	-	5	6	0.1490	-	3	5
2	0.0146	1.81	16	21	0.0260	1.74	9	14	0.0680	1.13	5	7
3	0.0041	1.85	30	34	0.0072	1.85	17	25	0.0248	1.46	7	11
4	0.0011	1.88	57	53	0.0019	1.92	31	40	0.0074	1.75	12	20
5	0.0003	1.90	98	81	0.0005	1.95	57	61	0.0020	1.89	22	34
6	7.9e-05	1.92	152	76	0.0001	1.96	99	85	0.0005	1.96	40	53
7	2.1e-05	1.92	215	72	3.2e-05	1.97	150	82	0.0001	1.99	71	76
8	6.1e-06	1.76	220	49	8.4e-06	1.95	190	58	3.2e-05	2.00	108	89
Lev/ ε	10^{-4}				10^{-5}				10^{-6}			
	Error	Order	(a)	(b)	Error	Order	(a)	(b)	Error	Order	(a)	(b)
1	0.1820	-	3	3	0.1920	-	2	2	0.1960	-	2	2
2	0.1090	0.74	3	5	0.1250	0.62	2	4	0.1290	0.60	2	4
3	0.0605	0.85	4	6	0.0820	0.61	3	5	0.0889	0.54	2	5
4	0.0270	1.16	5	8	0.0505	0.70	3	5	0.0607	0.55	2	5
5	0.0092	1.56	7	13	0.0267	0.92	4	7	0.0395	0.62	3	5
6	0.0026	1.81	14	25	0.0108	1.30	5	9	0.0234	0.76	3	6
7	0.0007	1.92	25	41	0.0034	1.66	8	17	0.0114	1.04	4	8
8	0.0002	1.98	45	62	0.0009	1.86	16	30	0.0042	1.44	5	10

Table 1: Uniform, $\mathcal{M}_h = R_h^{\text{orth}}BV_h$ (a)It-sBPV (b)It-mgGS

Tables 2 and 3 display results using Shishkin type meshes and the sBVP preconditioner along with both types of trial spaces outlined in Section 5. Here, N is related to the level according to $\text{Level} = \log_2(N) - 1$. According to [29, 31], standard Galerkin methods for (1.2) lead to a convergence rate of $\mathcal{O}(N^{-1} \ln N)$ using piecewise linear approximation. As shown in Tables 2 and 3, we obtain a convergence rate of $\mathcal{O}((N^{-1} \ln N)^2)$ using the SPLS method. The numerical tests appear to show that in the case of Shishkin meshes, the sBVP preconditioner appears to be robust with respect to ε for a fixed stopping criteria. In addition, one notable advantage of the preconditioner outlined in this paper is ease of implementation. All of the required parameters and matrices needed to implement the sBVP preconditioner, given in (8.3) or (8.4), are naturally computed in a standard implementation aside from the matrix that relates bases between spaces.

We mention here that, for the sBPV preconditioner on a Shishkin mesh, we took advantage of the fact that the Shishkin mesh is topologically equivalent

with a uniform mesh. We used a piecewise linear bijection (in each direction) to shift the uniform nodes to the Shishkin nodes. For the implementation of sBPV on Shishkin meshes, we used the same extension and restriction operators as in the case of uniform refinement. We plan to investigate the convergence of the new sBVP preconditioner on nonuniform refinements with suitable extension and restriction operators in a future work. This seems to be a challenging problem by itself and, by the best knowledge of the authors, has not been addressed in the case of Shishkin refinements or more general cases of fitted meshes.

Remark 6.1. *All numerical tests performed with the sBVP preconditioner used the preconditioner given by (8.3) as it results in a lower iteration count. At the expense of a slight increase in iteration count, the form given by (8.4) can be implemented to reduce the computational time.*

Lev/ ε	10^{-4}			10^{-6}			10^{-8}		
	Error	Order	It	Error	Order	It	Error	Order	It
1	0.2030	-	4	0.1970	-	3	0.1970	-	2
2	0.1570	0.89	5	0.1540	0.86	5	0.1540	0.86	4
3	0.1020	1.06	7	0.1020	1.03	6	0.1020	1.03	6
4	0.0539	1.36	11	0.0538	1.35	10	0.0538	1.35	9
5	0.0226	1.70	18	0.0226	1.70	17	0.0226	1.70	15
6	0.0079	1.94	30	0.0079	1.95	30	0.0079	1.95	28
7	0.0025	2.04	50	0.0025	2.04	55	0.0025	2.04	51
8	0.0008	2.05	88	0.0008	2.05	101	0.0008	2.05	95
Lev/ ε	10^{-10}			10^{-12}			10^{-14}		
	Error	Order	It	Error	Order	It	Error	Order	It
1	0.1970	-	3	0.1970	-	3	0.1970	-	2
2	0.1540	0.86	7	0.1540	0.86	7	0.1540	0.86	6
3	0.1020	1.03	11	0.1020	1.03	10	0.1020	1.03	9
4	0.0538	1.35	17	0.0538	1.35	16	0.0538	1.35	15
5	0.0226	1.70	28	0.0226	1.70	27	0.0226	1.70	25
6	0.0079	1.95	50	0.0079	1.95	47	0.0079	1.95	45
7	0.0025	2.04	90	0.0025	2.04	86	0.0025	2.04	83
8	0.0008	2.05	166	0.0008	2.05	159	0.0008	2.05	153

Table 2: Shishkin, $\mathcal{M}_h = R_h^{\text{orth}} B V_h$

7. CONCLUSION

We presented a preconditioning technique for the singularly perturbed reaction diffusion problem. We considered the concept of saddle point reformulation of the problem and the concept of optimal test norm as presented in [24, 27]. We showed the performance of our approach on a combination of

Lev/ ε	10^{-4}			10^{-6}			10^{-8}		
	Error	Order	It	Error	Order	It	Error	Order	It
1	0.2200	-	3	0.2140	-	3	0.2130	-	2
2	0.1800	0.70	4	0.1790	0.61	4	0.1790	0.60	4
3	0.1280	0.84	6	0.1280	0.84	6	0.1280	0.84	5
4	0.0776	1.07	8	0.0776	1.06	8	0.0777	1.06	7
5	0.0370	1.45	13	0.0369	1.45	12	0.0370	1.45	11
6	0.0138	1.84	24	0.0137	1.84	21	0.0137	1.84	20
7	0.0044	2.05	48	0.0043	2.06	39	0.0043	2.06	36
8	0.0013	2.08	100	0.0013	2.09	71	0.0013	2.09	67
Lev/ ε	10^{-10}			10^{-12}			10^{-14}		
	Error	Order	It	Error	Order	It	Error	Order	It
1	0.2130	-	3	0.2130	-	3	0.2130	-	2
2	0.1790	0.60	6	0.1790	0.60	6	0.1790	0.60	6
3	0.1280	0.84	9	0.1280	0.84	8	0.1280	0.84	8
4	0.0777	1.06	13	0.0777	1.06	12	0.0777	1.06	11
5	0.0370	1.45	21	0.0370	1.45	19	0.0370	1.45	18
6	0.0137	1.84	35	0.0137	1.84	34	0.0137	1.84	32
7	0.0043	2.06	64	0.0043	2.06	116	0.0043	2.06	58
8	0.0013	2.09	118	0.0013	2.09	113	0.0013	2.09	108

Table 3: Shishkin, $\mathcal{M}_h = R_h^{\text{lump}} BV_h$

two projection trial spaces and two different preconditioners that efficiently cover a wide range of the parameter ε . The method is also robust with respect to ε . The efficiency of the Uzawa preconditioned CG solver depends on the robustness and efficiency of preconditioners for the discrete *optimal norm* on the test space V_h . For quasi-uniform meshes, we introduced a simplified version of the Bramble-Pasciak-Vassilevski preconditioner. The numerical experiments demonstrate the preconditioner performs well even in the case of Shishkin type refinements. In the case of Shishkin meshes, we obtain higher order approximation of the gradient of the solution.

8. APPENDIX: A NOTE ON SBVP IMPLEMENTATION

Using the (\cdot, \cdot) inner product on V_h , we can identify V_h^* with V_h . The implementation of the sBVP preconditioner defined in (2.6) is done by computing the coordinate vector of the action of B_J^{sBVP} on dual vectors. To be more precise, let M_k be the mass matrix for the basis $\{\varphi_i^k, i = \overline{1 : n_k}\}$, and let D_k be the diagonal matrix with entries $D_{ii} = (1, \varphi_i^k)$. We define E_k to be the $n_J \times n_k$ matrix that relates the bases on V_k and V_J . That is, for the bases $\{\varphi_i^k, i = \overline{1 : n_k}\}$ of V_k and $\{\varphi_i^J, i = \overline{1 : n_J}\}$ of V_J , we have

$$(8.1) \quad (\varphi_1^k, \varphi_2^k, \dots, \varphi_{n_k}^k)^T = E_k^T (\varphi_1^J, \varphi_2^J, \dots, \varphi_{n_J}^J)^T.$$

For $f_h \in V_h$, the (dual) vector in \mathbb{R}^{n_k} is defined by

$$\underset{\sim_k}{f_h} := ((f_h, \varphi_1^k), (f_h, \varphi_2^k), \dots, (f_h, \varphi_{n_k}^k))^T.$$

For $w_h = \sum_{i=1}^{n_k} \alpha_i^k \varphi_i^k \in V_k \subset V_h$, the coordinate vector in \mathbb{R}^{n_k} is denoted by

$$\underset{\sim_k}{w_h} := (\alpha_1^k, \alpha_2^k, \dots, \alpha_{n_k}^k)^T.$$

Using (8.1) it is easy to check that

$$(8.2) \quad \underset{\sim_k}{f_h} = E_k^T \underset{\sim_J}{f_h}, \text{ and } \underset{\sim_J}{w_h} = E_k \underset{\sim_k}{w_h},$$

and by letting $w_h := \tilde{Q}_k f_h$, we have, due to (2.3) and (8.2),

$$\underset{\sim_J}{w_h} = E_k D_k^{-1} E_k^T \underset{\sim_J}{f_h}.$$

To obtain the contribution of $\gamma_J I$ in (2.6), we note that $\underset{\sim_J}{f_h} = M_J^{-1} \underset{\sim_J}{f_h}$.

Thus, the matrix version of (2.6) is given by

$$(8.3) \quad B_J^{sBVP} \underset{\sim}{f_J} = \gamma_J M_J^{-1} \underset{\sim}{f_J} + \sum_{j=0}^{J-1} (\gamma_j - \gamma_{j+1}) E_j D_j^{-1} E_j^T \underset{\sim}{f_J}.$$

To avoid mass matrix inversion, we can use the equivalence (2.7) and further simplify *sBVP* to the (matrix) version

$$(8.4) \quad B_J^{sBVP} \underset{\sim}{f_J} = \gamma_J D_J^{-1} \underset{\sim}{f_J} + \sum_{j=0}^{J-1} (\gamma_j - \gamma_{j+1}) E_j D_j^{-1} E_j^T \underset{\sim}{f_J}.$$

This way, we avoid mass matrix inversion and the iterative process is faster. We note that the matrix version of (2.5) just uses M_j instead of D_j in (8.4).

Acknowledgment. The first author is grateful for an Oden Institute fellowship which allowed him to visit UT Austin in the Fall of 2019.

REFERENCES

- [1] J. Adler, S MacLachlan, and N. Madden. A first-order system Petrov-Galerkin discretization for a reaction-diffusion problem on a fitted mesh. *IMA J. Numer. Anal.*, 36:1281–1309, 2016.
- [2] T. Apel. *Anisotropic finite elements: local estimates and applications*. Advances in Numerical Mathematics. B. G. Teubner, Stuttgart, 1999.
- [3] A. Aziz and I. Babuška. Survey lectures on mathematical foundations of the finite element method. *The Mathematical Foundations of the Finite Element Method with Applications to Partial Differential Equations*, A. Aziz, editor, 1972.
- [4] C. Bacuta. Schur complements on Hilbert spaces and saddle point systems. *J. Comput. Appl. Math.*, 225(2):581–593, 2009.
- [5] C. Bacuta, D. Hayes, and J. Jacavague. Notes on a saddle point reformulation of mixed variational problems. *Computers & Mathematics with Applications*, 2020.

- [6] C. Bacuta and J. Jacavage. Saddle point least squares for the reaction diffusion problem. *Results in Applied Mathematics*, Available online 7 May 2020:1–20, 2019.
- [7] C. Bacuta and J. Jacavage. Saddle point least squares preconditioning of mixed methods. *Computers & Mathematics with Applications*, 77(5):1396–1407, 2019.
- [8] C. Bacuta and J. Jacavage. Least squares preconditioning for mixed methods with nonconforming trial spaces. *Applicable Analysis*, Available online Feb 27, 2019:1–20, 2020.
- [9] C. Bacuta and J. Jacavage. A non-conforming saddle point least squares approach for an elliptic interface problem. *Computational Methods in Applied Mathematics*, doi:10.1515/cmam-2018-0202, 2019.
- [10] C. Bacuta and P. Monk. Multilevel discretization of symmetric saddle point systems without the discrete LBB condition. *Appl. Numer. Math.*, 62(6):667–681, 2012.
- [11] C. Bacuta and K. Qirko. A saddle point least squares approach to mixed methods. *Comput. Math. Appl.*, 70(12):2920–2932, 2015.
- [12] C. Bacuta and K. Qirko. A saddle point least squares approach for primal mixed formulations of second order PDEs. *Comput. Math. Appl.*, 73(2):173–186, 2017.
- [13] T. Bouma, J. Gopalakrishnan, and A. Harb. Convergence rates of the DPG method with reduced test space degree. *Comput. Math. Appl.*, 68(11):1550–1561, 2014.
- [14] J.H. Bramble, J.E. Pasciak, and P. Vassilevski. Computational scales of Sobolev norms with application to preconditioning. *Math. Comp.*, 69(230):463–480, 2000.
- [15] S. Brenner and L.R. Scott. *The mathematical theory of finite element methods*. Springer-Verlag, New York, 1994.
- [16] D. Broersen and R. Stevenson. A robust Petrov-Galerkin discretisation of convection-diffusion equations. *Comput. Math. Appl.*, 68(11):1605–1618, 2014.
- [17] L. Demkowicz C. Carstensen and J. Gopalakrishnan. Breaking spaces and form for the DPG method and applications including maxwell equations. *Computers and Mathematics with Applications*, 72:494–522, 2016.
- [18] Z. Cai and J. Ku. A dual finite element method for a singularly perturbed reaction-diffusion problem. *SIAM J. Numer. Anal.*, 58(3):1654–1673, 2020.
- [19] 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(4):771–795, 2014.
- [20] A. Cohen, W. Dahmen, and G. Welper. Adaptivity and variational stabilization for convection-diffusion equations. *ESAIM Math. Model. Numer. Anal.*, 46(5):1247–1273, 2012.
- [21] L. Demkowicz and J. Gopalakrishnan. A class of discontinuous Petrov-Galerkin methods. Part I: the transport equation. *Comput. Methods Appl. Mech. Engrg.*, 199(23-24):1558–1572, 2010.
- [22] L. Demkowicz and J. Gopalakrishnan. A class of discontinuous Petrov-Galerkin methods. ii. optimal test functions. *Numerical Methods for Partial Differential Equations*, 27(1):70–105, 2011.
- [23] L. Demkowicz and L. Vardapetyan. Modelling electromagnetic/scattering problems using hp-adaptive finite element methods. *Comput. Methods Appl. Mech. Engrg. Numerical Mathematics*, 152:103 – 124, 1998.
- [24] J Gopalakrishnan. Five lectures on DPG methods. *arXiv 1306.0557*, 2013.
- [25] K. W. Morton J. W. Barrett, and. Optimal Petrov-Galerkin methods through approximate symmetrization. *IMA J. Numer. Anal.*, 1(4):439–468, 1981.
- [26] T. Kolev, J. Xu, and Y. Zhu. Multilevel preconditioners for reaction-diffusion problems with discontinuous coefficients. *J. Sci. Comput.*, 67(1):324–350, 2016.
- [27] N. Heuer L. Demkowicz, T. Führer and X. Tian. The Double Adaptivity Paradigm (how to circumvent the discrete inf-sup conditions of Babuška and Brezzi). Technical report, Oden Institue for Computational Engineering and Sciences, The University fo Texas at Austin, May 2019.

- [28] J. Li. Convergence and superconvergence analysis of finite element methods on highly nonuniform anisotropic meshes for singularly perturbed reaction–diffusion problems. *Applied Numerical Mathematics*, 36(2):129–154, 2001.
- [29] R. Lin and M. Stynes. A balanced finite element method for singularly perturbed reaction-diffusion problems. *SIAM Journal on Numerical Analysis*, 50(5):2729–2743, 2012.
- [30] S. MacLachlan and N. Madden. Robust Solution of Singularly Perturbed Problems using Multigrid Methods. *SIAM J. Sci. Comput.*, 35(5):2225–2254, 2013.
- [31] H.G. Roos and M. Schopf. Convergence and stability in balanced norms of finite element methods on shishkin meshes for reaction-diffusion problems: Convergence and stability in balanced norms. *ZAMM Journal of applied mathematics and mechanics: Zeitschrift für angewandte Mathematik und Mechanik*, 95(6):551–565, 2014.
- [32] G.I. Shishkin. Grid approximation of singularly perturbed boundary value problems with a regular boundary layer. *Sov. J. Numer. Anal. Math. Model.*, 4(5):397–417, 1989.

UNIVERSITY OF DELAWARE, MATHEMATICAL SCIENCES, 501 EWING HALL, NEWARK,
DE 19716

Email address: `bacuta@udel.edu`

UNIVERSITY OF DELAWARE, DEPARTMENT OF MATHEMATICS, 501 EWING HALL 19716

Email address: `dphayes@udel.edu`

LAFAYETTE COLLEGE, DEPARTMENT OF MATHEMATICS, PARDEE HALL, EASTON, PA
18042

Email address: `jacavagj@lafayette.edu`